On the characterisation of the underlying event

Matteo Cacciari,[1,2] Gavin P. Salam[1] and Sebastian Sapeta[1]

[1]*LPTHE, UPMC Univ. Paris 6 and CNRS UMR 7589, Paris, France*
[2]*Université Paris Diderot, Paris, France*
E-mail: `cacciari@lpthe.jussieu.fr, salam@lpthe.jussieu.fr,`
`sapeta@lpthe.jussieu.fr`

Abstract

The measurement of the underlying event (UE) and its separation from hard interactions in hadron-collider events is a conceptually and practically challenging task. We develop a simple, mostly analytical toy model for the UE in order to understand how different UE measurement approaches fare on the practical aspects of this problem, comparing the traditional approach used so far at Tevatron with a recently proposed "jet-area/median" approach. Both are found to perform comparably well in measuring average properties of the UE, such as the mean transverse momentum flow, but the jet-area/median approach has distinct advantages in determining its fluctuations. We then use the jet-area/median method to investigate a range of UE properties in existing Monte Carlo event-generator tunes, validating the main results of the toy-model and highlighting so-far unmeasured characteristics of the UE such as its rapidity dependence, as well as its intra- and inter-event fluctuations and correlations.

Contents

1 Introduction

The "underlying event" (UE) in high-energy hadron-hadron collisions can be thought of as the low transverse momentum (p_t) part of the event activity that is not naturally associated with the hard interaction.

Despite being a low-p_t phenomenon, the underlying event has a large impact on high-p_t physics at hadron colliders. For example in measurements of the inclusive jet spectrum at $p_t \sim 50\,\text{GeV}$ it can affect the result by up to 50% [1]. It also biases kinematic reconstructions (for example in top mass measurements), degrades their resolution and affects the efficiency of isolation criteria that enter into the experimental identification of particles like photons and electrons. It has even been suggested that certain new physics scenarios might show up in "anomalous" characteristics of the underlying event [2]. For these reasons it is important to have a good understanding of its properties.
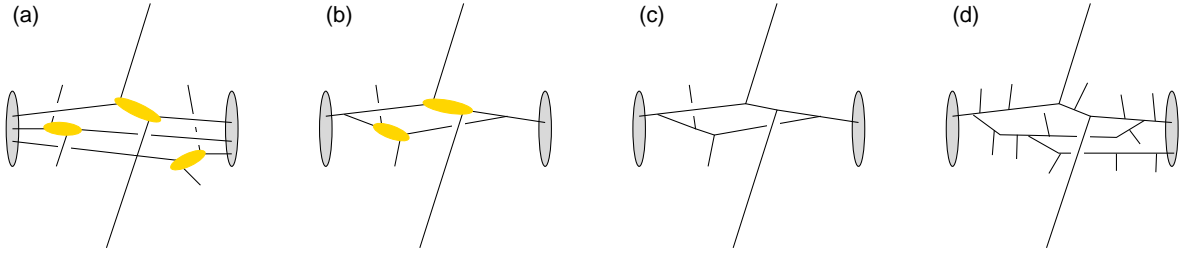
Figure 1: Views of a hard scattering plus UE; (a) shows multiple $2 \to 2$ interactions, as incorporated in the most successful models; (b) illustrates how a collinear splitting in the initial-state can lead to correlation between the partons involved in a double $2 \to 2$ scattering; from (c) one sees the similarity with a perturbative 1-loop $2 \to 4$ diagram; and (d) represents a BFKL-inspired picture for the UE.

The purpose of this article is to investigate some different ways in which the underlying event can be measured and/or constrained experimentally. Such measurements enter into tunes of Monte Carlo models of the UE [3–10]. They can also serve as an input to analytical methods of accounting for the average UE correction to a jet's transverse momentum [11, 12] and to approaches that correct for the UE on a jet-by-jet and event-by-event basis [13], as well as for related work that seeks to optimise jet definitions.

A difficulty in discussing the measurement of the UE is that there exists no good definition of what the UE actually is, or how to distinguish it, in a conceptually unambiguous way, from the hard interaction. For instance, the most successful phenomenological models of the UE involve multiple $2 \to 2$ scattering as in fig. 1a ("multiple parton interactions" — MPI). Very simply, they supplement the one hard interaction in the event with multiple other lower-p_t interactions, whose multiplicity is determined by the $2 \to 2$ cross section, regulated with a low-p_t cutoff of the order of a couple of GeV. Fig. 1a can only be part of the picture because some of the partons entering multiple $2 \to 2$ scatterings are necessarily correlated, e.g. due to energy conservation (cf. ref. [14]), or because they can have a common origin from an initial-state collinear splitting, as in fig. 1b. However the contribution of fig. 1b is itself also part of the $2 \to 4$ 1-loop scattering diagram fig. 1c (as discussed for example in [15]), which is relevant as a N^{4}LO correction to the dijet cross section or at N^{3}LO in its interference with tree-level $2 \to 4$ scattering. This means there are non-trivial questions of double-counting between multiple parton interaction and perturbative higher orders. In addition, the radiation that fills the event is not bound to come just from $2 \to 2$ scatterings, but may also arise from BFKL type configurations which can involve (multiple) chains with low-p_t emissions spread in rapidity, as in fig. 1d (some work towards modelling this while retaining consistency with the total cross section is to be found in [16]). Though these ways of viewing the UE represent just a subset of the diversity of physics considerations that are of potential relevance for its modelling (more detailed reviews are to be found in Refs. [10, 17]), they do illustrate the difficulties that arise in ascribing unambiguous physical meaning to it.

Given this complexity in discussing what the UE might be, how are we to go about measuring it? A feature present in most models is that, on some low p_t scale, UE activity fills the whole event. One way then of characterising the UE is to say that it *is* whatever physical effect fills most of the event with radiation. To help understand what implications this picture has for UE measurements, we shall develop (section 3) a semi-analytical two-component toy

model: one component will be purely soft and dispersed across the event, corresponding to the UE, while the second component will involve the hard scattering and a simple approximation for the perturbative radiation with which it is associated.

Though our toy model is undoubtedly too simple to fully reflect reality, the fact that we know exactly what goes into it will make it quite powerful: it will, for example, allow us to take different UE-measurement approaches and examine to what extent their results are affected both by the radiation associated with the hard scattering, and by the techniques used to limit that hard contamination. This can be investigated both for averaged quantities, and for event-by-event extractions of information about the UE. A number of the results will be given in analytical form, in terms of the characteristics of the UE and the hard scattering and of the parameters of the measurement methods. This will give insight into the compromises that arise when measuring the UE, especially when extending existing methods to the greater phase space that is available at LHC relative to Tevatron.

The two UE measurement methods that we shall investigate are both reviewed in section 2. One, which we call the "traditional" approach (see for example [18,19]), is currently the default approach for most UE studies. Another, the jet-area/median based approach of [12,13], was originally developed for evaluating pp pileup (and backgrounds in heavy-ion collisions, as used for example in [20]), but may also have benefits for UE studies. The basic results for the two approaches will be derived in section 3. Some readers may prefer to skip most of this section and read just the final summary of these results, as presented in section 3.6.

For the purpose of determining the quantity that we call ρ, the UE's mean transverse momentum (p_t) flow per unit rapidity-azimuth area, both methods will turn out to have systematics that are under control and of similar magnitude, at about the 20% level (except at high p_t for the traditional method). However it is also important to have knowledge of fluctuations of the underlying event (both intra and inter-event). Since it is the jet-area/median method that will prove to be the more robust tool for measuring them, it is this method that we will use when, in section 4, we look at a range of possibly interesting measurements of the UE. They will be carried out on Monte Carlo events simulated with Pythia 6.4 [21] and with Herwig 6.5 [22] with Jimmy 4.3 [4]. This part of the study will help validate the understanding developed with the toy-model, and illustrate determinations of the UE average p_t flow as a function of rapidity, its event-to-event fluctuations, its intra-event fluctuations, and the degree of intra-event correlation. These observables go beyond the kinds of measurements that are commonly discussed for Tevatron or envisaged so far for the LHC and, as we shall see, there will be substantial differences between Jimmy and Pythia UE tunes on a number of them.

2 Overview of measurement approaches

2.1 Traditional approach

The main current approach [18] to measuring the properties of the underlying event involves considering the central part of the detector, say $|\eta| < 1$, where the pseudorapidity η is defined as $\eta = -\ln\tan\frac{\theta}{2}$. One tags events based on the presence of a jet (whose direction defines an azimuthal angle $\phi = 0$), and divides the central part of the event into four blocks in azimuth: the "towards" region, typically $|\phi| < \pi/3$, an away region $2\pi/3 < |\phi| < \pi$, and two transverse regions, covering $\pi/3 < |\phi| < 2\pi/3$. This is illustrated in fig. 2 (left). Since the trigger and recoil jets will usually occupy the towards and away regions, one then restricts one's attention
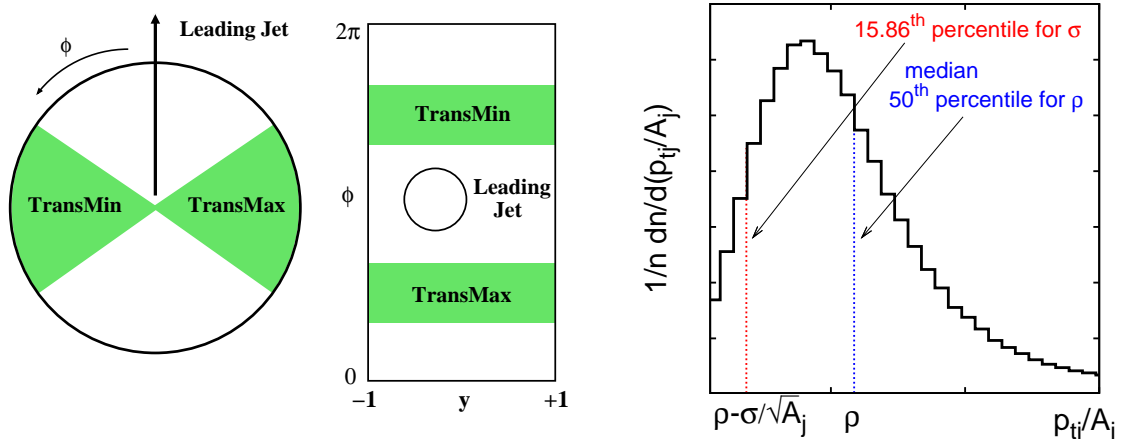
Figure 2: Illustration of underlying-event measurement methods. Left: representation of TransMin and TransMax regions of the rapidity azimuth plane in the traditional approach. Right: representative distribution of p_{tj}/A_j for the set of (real and ghost) jets in a single event, as used to determine ρ and σ in the area/median approach.

to the two transverse regions. There one measures the multiplicity of charged tracks above some transverse-momentum threshold as well as the total transverse momentum contained in the charged tracks (sometimes normalised per unit area, $d\eta d\phi$). The results for the charged track multiplicity and charged p_t flow are usually presented as averages across many events, as a function of the p_t of the leading jet. One also sees measurements of the charged momentum flow as a function of the multiplicity.

Since there is a probability of order α_s that at least one of the transverse regions is contaminated by perturbative radiation from the dijet event, which substantially affect the extracted information about the UE's p_t flow. To work around this, it is usual to label the two transverse regions as TransMin and TransMax, respectively the less and more active of the two. The largest component of perturbative contamination should be restricted to the TransMax region, while TransMin should be less affected.

In the earliest variants of this "two-region" method [23, 24], the two regions used for sampling the UE were actually placed at non-central rapidities, rather than central rapidities and transverse azimuth, and it was the total transverse energy flow that was considered rather than just its charged component. Another variant measured charged momentum flow in cones [25]. These differences reflect the freedom inherent to this method: the question of where to place the "transverse" regions, and the choice of their shape and size. In the rest of this article we will always assume that the transverse regions are well separated from the dijet system in an event, as shown in fig. 2 (left), and we will quote our results as a function of the *area* A_{Trans} of each of the transverse regions.

2.2 Jet-area/median approach

In the jet-area/median approach, one first clusters the event with a Cambridge/Aachen (C/A) [26, 27] or inclusive k_t [28, 29] type jet algorithm. To each jet, j, one attributes an "active" jet-area, A_j, as described in more detail in [12]. This is calculated by adding a large number of "ghost" particles to the event (each with negligible $p_t \sim 10^{-100}$ GeV) and including them

4

in the clustering. The area of a jet is then proportional to number of ghosts it contains. Some jets contain just ghost particles ("pure ghost jets") and are considered to have $p_t = 0$.

A proposal [13] for a way to measure the level ρ of uniform background noise in an event is then to take it to be the median of the distribution of the p_{tj}/A_j for the ensemble of jets in that event

$$\rho = \underset{j \in \text{jets}}{\text{median}} \left[\left\{ \frac{p_{tj}}{A_j} \right\} \right] , \tag{1}$$

as shown schematically in fig. 2 (right). The logic of the use of the median is that it is much less susceptible to contamination from outliers (i.e. hard perturbative jets) than the mean. In addition to measuring ρ one can also determine the intra-event fluctuations of the UE. We introduce a quantity σ, defined such that a fraction $X/2$ of jets satisfy

$$\rho - \sigma/\sqrt{\langle A_j \rangle} < p_{tj}/A_j < \rho , \tag{2}$$

where $X = \text{Erf}(1/\sqrt{2}) \simeq 0.6827$ is the fraction of a Gaussian distribution within one standard deviation of the mean. [1] The approach to measuring σ is analogous in spirit to the use of the median for determining ρ. As shown in fig. 2 (right), it is one sided (i.e. just considering jets with $p_{tj}/A_j < \rho$). This choice has been made so as to limit contamination of σ from hard jets when the total number of jets is small.

This method was originally suggested in [13] as a way of measuring average pileup noise across an entire event (e.g. for $|y| < 5$). One difficulty when using it for the UE will come from the fact that the UE is significantly softer, and has relative fluctuations that are larger and less Gaussian than those in (say) 20 pileup minimum-bias events, both consequences of the lower density of particles.

Another point is that in measuring the UE it is important to obtain differential information on the UE's rapidity ($y = \ln \frac{E+p_z}{E-p_z}$) dependence. This leads us to consider

$$\rho(y) = \underset{j \in \text{jets}, \, |y_j - y| < \delta y}{\text{median}} \left[\left\{ \frac{p_{tj}}{A_j} \right\} \right] , \tag{3}$$

where $2\delta y$ is the width of a rapidity window in which one carries out the measurement. The fact that one has a relatively limited number of jets in a given rapidity window will be one further challenge that we will face in using this method for studying the UE, because the relative impact of the presence of a hard jet in the region of interest is amplified by the small total number of jets. A modification of the method that helps address these difficulties is "hard jet removal," first employed in the STAR collaboration's [30] use of the techniques of [13] for estimating the (very large) UE for jet measurements in heavy-ion collisions. We will investigate the impact of this modification here for events with dijet topologies, in which case we will simply remove the two hardest jets in the event from the overall list of jets.

A choice that is present in both pileup and UE measurements with the area/median method is that of the jet algorithm and jet radius R. Both the k_t [28, 29] and Cambridge/Aachen [26, 27] algorithms are suitable options, because they produce jets whose area distribution is quite regular. In contrast, algorithms that give mostly conical jets (like anti-k_t [31] and, to a lesser extent, SISCone [32]) tend not to be, because they fill in the "holes"

---

[1] This definition is such that, if the noise distribution is Gaussian and uncorrelated from point to point within the event, then a patch of area A will have an average noise p_t contamination of ρA and standard deviation $\sigma \sqrt{A}$.

between the cones with jets with very small areas, which can have unrepresentative p_t/A values. The question of what R value to use is one of the freedoms of the method and will be discussed in the coming section.

3   A toy model

To understand the strengths and weaknesses of different UE-measurement approaches, it is helpful to consider events as consisting of two components: a low-p_t ("soft") noise component, *defined* to be the UE, supplemented with hard jets from a perturbative scattering and associated higher-order corrections. We will introduce models for each of these two components and investigate how the methods behave when either of the two components is present alone and when both are present together. Our guiding principle in designing these models has been to keep them sufficiently simple as to be treatable analytically, while also maintaining a reasonable degree of realism. While our combined hard and soft models will not quite have the continuous transition between hard and soft components that is present in Monte Carlo MPI models, we will see that they nevertheless lead to certain signature behaviours in UE measurement methods that correspond nicely to what is observed in Monte Carlo simulations.

3.1   Low-p_t component

As a simple model for the underlying event let us imagine that on average in a patch of unit of area there are ν particles, that the probability distribution for the number of particles n in a specific patch of area A follows a Poisson distribution, $P_n = (\nu A)^n e^{-\nu A}/n!$, and that the single-particle transverse-momentum probability distribution is given by

$$\frac{1}{P_1}\frac{dP_1}{dp_t} = \frac{1}{\mu}e^{-p_t/\mu}, \tag{4}$$

where μ is the mean transverse momentum per particle. This particular form has been chosen mainly because it will allow us to carry out analytical calculations.[2] If a patch contains n particles then the probability distribution for its p_t is given by

$$\frac{1}{P_n}\frac{dP_n}{dp_t} = \frac{1}{(n-1)!}\frac{p_t^{n-1}}{\mu^n}e^{-p_t/\mu}. \tag{5}$$

For a patch of area A, summing over the Poisson distribution for the number of particles in the patch ($\langle n \rangle = \nu A$) gives us the overall probability distribution for the transverse momentum in the patch as

$$\frac{dP}{dp_t}(A) = \sum_{k=0}^{\infty}\frac{dP_k}{dp_t} = \delta(p_t)e^{-\nu A} + \sum_{n=1}^{\infty}\frac{(\nu A)^n}{n!}e^{-\nu A}\frac{1}{(n-1)!}\frac{p_t^{n-1}}{\mu^n}e^{-p_t/\mu} \tag{6a}$$

$$= \delta(p_t)e^{-\nu A} + e^{-\nu A - p_t/\mu}\sqrt{\frac{A\nu}{\mu p_t}}\,I_1\left(2\sqrt{\frac{A\nu p_t}{\mu}}\right), \tag{6b}$$

---

[2]Another simple variant will be considered in appendix A. It would also be interesting to consider a distribution suppressed as $1/p_t^n$ at large p_t, more in line with the scaling that is to be found with multiple parton interactions. However, we have not found a form with this property that can be handled analytically throughout the calculation. Given that the results from the toy model will reproduce many features that we see in Monte Carlo simulation, we believe that the form we are using is adequate for our purposes.

where I_1 is the (first order) modified Bessel function of the first kind. The mean and standard deviation of the distribution are given by $\nu A\mu$ and $\sqrt{2\nu A}\mu$. It is convenient to express this as saying that the transverse momentum in a patch of area A is

$$p_t(A) = \rho A \pm \sigma \sqrt{A}, \tag{7}$$

where, in the model discussed here, ρ and σ are given by

$$\rho = \nu\mu, \tag{8a}$$

$$\sigma = \sqrt{2\nu}\mu = \rho\sqrt{\frac{2}{\nu}}. \tag{8b}$$

In the limit in which $\nu A \gg 1$, the distribution in eq. (6) tends to a Gaussian with mean and standard deviation as given in eqs. (7) and (8). This is a consequence of the central-limit theorem.

Traditional approach. Taking the area of each of the transverse regions to be A_{Trans}, the traditional approach will extract the following results for ρ in the transverse Average, Min and Max regions

$$\langle \rho_{\text{ext,Av}} \rangle = \rho, \tag{9a}$$

$$\langle \rho_{\text{ext,Min}} \rangle = \rho - \frac{\sigma}{\sqrt{\pi A_{\text{Trans}}}}, \tag{9b}$$

$$\langle \rho_{\text{ext,Max}} \rangle = \rho + \frac{\sigma}{\sqrt{\pi A_{\text{Trans}}}}, \tag{9c}$$

where the Min and Max results have been derived using the Gaussian limit of eq. (6). The only one of the above results that correctly estimates ρ is the average. The Min and Max results tend slowly to the correct answer in the limit in which A_{Trans} is large.

In the literature (e.g. [18]), A_{Trans} has usually been taken equal to $2\pi/3$. As we shall see in section 4, a typical value for σ at the LHC ($\sqrt{s} = 10\,\text{TeV}$) is $\sigma \simeq 0.5\rho - 0.75\rho$, which implies that the $\sigma/\sqrt{\pi A_{\text{Trans}}}$ term is about $20 - 30\%$ of ρ.

Area/median-based approach. To help understand the behaviour of the area/median-based approach, let us replace the jets (which have a range of areas) with uniform rectangular tiles, each of which has a fixed area A_{tile}. It is important to use the full distribution $dP/dp_t(A_{\text{tile}})$ as given by eq. (6) rather than the Gaussian distribution, because a physically interesting domain is that in which νA_{tile} is of order 1. The extracted value ρ_{ext} of the UE p_t density in the tiled approximation is given by the median value of $p_{t,\text{tile}}/A_{\text{tile}}$ across the many tiles in a single event. It can be determined from the solution of the equation

$$\int_0^{A_{\text{tile}}\,\rho_{\text{ext}}} dp_t \frac{dP}{dp_t}(A_{\text{tile}}) = \frac{1}{2}, \tag{10}$$

This result has been obtained in the limit of there being a large number of tiles, i.e. large A_{tot}, which allows us to approximate the distribution of tile transverse momenta in a specific event with the average probability distribution $\frac{dP}{dp_t}(A_{\text{tile}})$ (see also appendix B.1). The integral in eq. (10) is non-trivial to evaluate analytically, however an approximation to the solution for

ρ_{ext} that is accurate to a couple of percent and has the correct asymptotic behaviours is given by

$$\rho_{\text{ext}} \simeq \rho \, \frac{\nu A_{\text{tile}} - \ln 2}{\nu A_{\text{tile}} - \ln 2 + \frac{1}{2}} \, \Theta(\nu A_{\text{tile}} - \ln 2) \,. \tag{11}$$

The result is non-zero only for $\nu A_{\text{tile}} > \ln 2$, which stems from the requirement that tiles with no particles, i.e. the $\delta(p_t)$ contribution in eq. (6), should not account for more than half of the total number of tiles. This property of a sudden turn-on, as well as the fact that at large νA_{tile} the offset from the correct ρ goes as $1/(\nu A_{\text{tile}})$,

$$\frac{\rho_{\text{ext}} - \rho}{\rho} = -\frac{1}{2\nu A_{\text{tile}}} + \mathcal{O}\left((\nu A_{\text{tile}})^{-2}\right) \,, \tag{12}$$

are features that we have found to hold for certain other analytic forms of dP_1/dp_t, notably all those with a structure $p_t^m e^{-(m+1)p_t/\mu}$ (for arbitrary positive m). Other characteristics, such as the particular coefficient of the $1/(\nu A_{\text{tile}})$ offset in eq. (12), or the analytic structure close to the turn-on, do depend on the form taken for dP_1/dp_t.

The determination of ρ from the median of jets' p_t densities differs from the above "tiled" model in that jets do not have a fixed area. There is no simple way of extending the analytical model so as to account for this, however one can study the impact of the question numerically, by examining toy events in which many soft particles have been generated according to eq. (4), with random positions in the $y - \phi$ plane. On each event one runs the jet-based procedure to determine ρ_{ext} and compares it to ρ_{ext} for the tile-based procedure and to the analytic approximation eq. (11). We assume that tiles of A_{tile} are comparable to jets of the same average area, $A_{\text{tile}} = \langle A_{\text{jet}} \rangle$. One might think that the average jet area should be given by the result for ghost jets in Ref. [12], $\langle A_{\text{ghost-jet}} \rangle \simeq 0.55\pi R^2$. However, this only holds in the limit of very dense UE; for the typical kinds of configuration that are of interest to us, it will be more appropriate to use

$$\langle A_{\text{jet}} \rangle \simeq c_J R^2 \,, \qquad c_J \simeq 0.65\pi \simeq 2.04 \,, \tag{13}$$

where we have defined a constant c_J that will reappear in several places below.

Given this relation between the typical jet area and radius R, we can deduce the critical radius, R_{crit}, below which ρ is zero,

$$R_{\text{crit}} \simeq \frac{\sigma}{\rho} \cdot \sqrt{\frac{\ln 2}{2c_J}} \simeq 0.41 \frac{\sigma}{\rho} \,. \tag{14}$$

We can also rewrite eq. (12) in terms of R and σ/ρ or R_{crit}:

$$\frac{\rho_{\text{ext}} - \rho}{\rho} = -\frac{\sigma^2}{4\rho^2 c_J R^2} + \mathcal{O}\left(\frac{\sigma^4}{\rho^4 R^4}\right) = -\frac{1}{2\ln 2} \frac{R_{\text{crit}}^2}{R^2} + \mathcal{O}\left(\frac{R_{\text{crit}}^4}{R^4}\right) \,. \tag{15}$$

The above results depend on the specific form of toy model that one chooses. To estimate the importance of this model dependence, one can replace dP_1/dp_t as given in eq. (4) with the alternative form

$$\frac{1}{P_1} \frac{dP_1}{dp_t} = 4p_t \mu^{-2} e^{-2p_t/\mu} \,, \tag{16}$$

as discussed in more detail in appendix A. The essential relations for this model are $\sigma/\rho = \sqrt{3/(2\nu)}$, $R_{\text{crit}} = \frac{\sigma}{\rho}\sqrt{(2\ln 2)/(3c_J)} \simeq 0.48\frac{\sigma}{\rho}$ and

$$\frac{\rho_{\text{ext}} - \rho}{\rho} = -\frac{2\sigma^2}{9\rho^2 c_J R^2} + \mathcal{O}\left(\frac{\sigma^4}{\rho^4 R^4}\right) = -\frac{1}{3\ln 2} \frac{R_{\text{crit}}^2}{R^2} + \mathcal{O}\left(\frac{R_{\text{crit}}^4}{R^4}\right) \,. \tag{17}$$
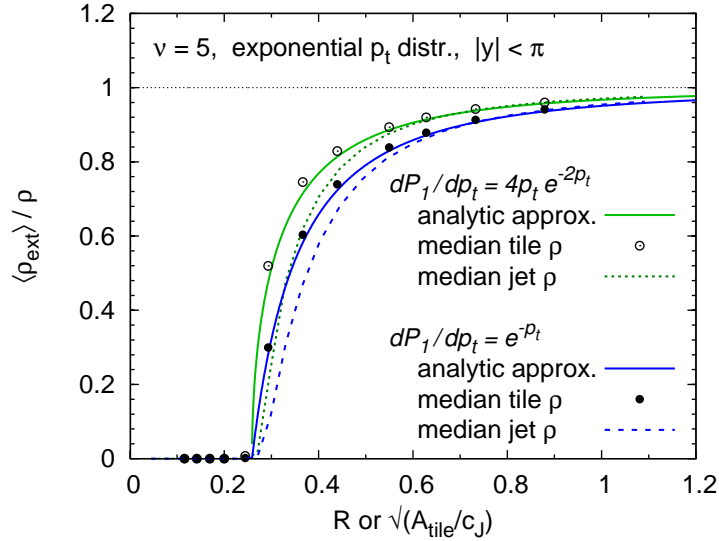
Figure 3: Comparison of jet and tile median-based determinations of ρ for two toy soft-event models (as described in the text), together with the corresponding analytical approximations. The mean number of particles per unit area, ν, was 5 (commensurate with typical expectations for the UE at LHC) and the jets that were used had $|y| < \pi$ (this choice stems from our use of square tiles). To achieve sub-percent-level agreement between the analytical and jet-based results at larger R values, it was necessary to carry out the determination of ρ_{ext} using p_t-scheme recombination of 4-momenta in the jets.

These results involve the same analytic structures as for the original form of the toy model, with numerical coefficients that imply slightly smaller corrections for finite R.

In fig. 3, the approximate analytical toy model results as a function of R (or the equivalent A_{tile}) are compared to the average results obtained by applying the "tiled" median approach as well as the jet-area median approach to toy-model configurations. This is done for both variants of the toy model. There is near perfect agreement between the analytical approximations and the average median tile-based results. This is indicative both of the quality of the analytical approximation and of the limited impact of the practical use of finite A_{tot} (here $4\pi^2$), as opposed to the large A_{tot} limit that went into the analytical results. The median jet-based results are rather close to the tile-based results for larger R values,[3] though the precise shape in the turn-on region differs a little with respect to the tile-based expectation. Moderate differences exist between the results with the two choices for dP_1/dp_t, and these can be taken as indicative of the magnitude of the model-dependence in the above analysis.

One message to take from fig. 3 is that for an R value that is twice that where the turn-on occurs, ρ_{ext} underestimates ρ by about $10 - 20\%$. This can be kept in mind as a ballpark value for the accuracies that we will be able to achieve and can be compared to the $20 - 30\%$ effect discussed above for the TransMin ρ determination in the traditional approach.

---

[3]To get this level of agreement for $R \sim 1$, it turned out to be necessary to use the p_t recombination scheme in determining the jet momenta.

3.2 Purely perturbative events

In our definition of purely perturbative events there is no underlying event and the *only* regions of the event that have non-zero momentum are those that contain a perturbative emission. Nevertheless UE-determination methods may still give a non-zero result for their estimate of the UE energy density, because of the way they are affected by those extra jets.

We work here with the assumption that we have selected events with at least two hard jets (or with a W or Z boson), and that extra jets may be present at higher perturbative orders. A crude, but illustrative approximation for those higher orders will be obtained as follows. We will take the dominant source of emissions to be radiation from the initial-state partons that enter the hard reaction.[4] Furthermore we will assume that those emissions are soft, distributed uniformly in rapidity and azimuth, and with a p_t spectrum given by $\alpha_s(p_t)/p_t$:

$$\frac{dn}{dp_t dy d\phi} \simeq \frac{C_i}{\pi^2}\frac{\alpha_s(p_t)}{p_t}\,, \tag{18}$$

where C_i is the colour factor associated with the incoming partons ($C_i = C_A \equiv 3$ for gluons, $C_i = C_F \equiv \frac{4}{3}$ for quarks). For a dijet event whose two hard jets have transverse momenta $p_{t,\text{hard}}$, we will take eq. (18) to be valid independently of rapidity for $Q_0 < p_t < Q$, where $Q \sim \frac{1}{2}p_{t,\text{hard}}$ and $Q_0 \sim 1\,\text{GeV}$.[5] We will also assume that emissions are independent. Thus, the probability distribution for the number of emissions will be a Poisson distribution whose mean is obtained by the integral of eq. (18) over the relevant phase space.

Traditional approach. The average p_t densities in the Average and the Max regions will both receive contributions from the emission of one gluon (relative to the Born diagram for the process). In contrast, the Min region only receives a contribution when at least two gluons have been emitted. One can obtain the $\rho_{\text{ext,Av}}$ value just by integrating eq. (18) up to Q, which we do in a fixed coupling approximation, $\alpha_s = \alpha_s(Q)$, since the integral is dominated by values of $p_t \sim Q$:

$$\langle \rho_{\text{ext,Av}} \rangle = \frac{1}{2A_{\text{Trans}}} \int_0^Q p_t dp_t \int_{2A_{\text{Trans}}} dy d\phi \, \frac{dn}{dp_t dy d\phi} = \frac{C_i \alpha_s}{\pi^2}Q\,. \tag{19}$$

Here we have neglected the (small) impact of Q_0 and, in the fixed-coupling approximation, the result is complete to all orders in α_s. One feature to note about the result is that it scales with Q.

To determine $\langle \rho_{\text{ext,Min}} \rangle$ to $\mathcal{O}\left(\alpha_s^2\right)$, we assume that the left and right transverse regions each contain one gluon, and that the left-hand gluon (L) is harder than the right-hand one (R); $\langle \rho_{\text{ext,Min}} \rangle$ is then given by the average transverse momentum in the right-hand region,

$$\langle \rho_{\text{ext,Min}} \rangle = 2\frac{1}{A_{\text{Trans}}}\left(\frac{C_i \alpha_s A_{\text{Trans}}}{\pi^2}\right)^2 \int_0^Q \frac{dp_{t,L}}{p_{t,L}} \int_0^{p_{t,L}} \frac{dp_{t,R}}{p_{t,R}}p_{t,R} = 2\left(\frac{C_i \alpha_s}{\pi^2}\right)^2 A_{\text{Trans}}\,Q\,, \tag{20}$$

---

[4]Ignoring radiation from any outgoing (Born) partons is not too poor an approximation, because a significant part of that radiation will be contained within the corresponding jets.

[5]Based on collinear factorisation, one would expect that the upper limit on the p_t of emissions to have significant rapidity dependence. For example, if the hard process takes place at central rapidities, then one might write $p_t \lesssim p_{t,\text{hard}}e^{-|y|}$. The rapidity-independent approximation is instead inspired by a high-energy factorisation picture, relevant when $p_t \ll \sqrt{s}e^{-|y|}$. Studies with Herwig at parton-level (based on collinear factorisation) give a distribution for the upper p_t limit on extra jets that is intermediate between these two expressions.

with an additional factor of 2 to account for the case where the right-hand gluon is the harder one. The result is proportional to α_s^2, i.e. suppressed by an extra factor of α_s compared to $\rho_{\text{ext,Av}}$, however it is enhanced by a factor of A_{Trans}.

Finally we can estimate $\langle \rho_{\text{ext,Max}} \rangle$ using the relation $\rho_{\text{ext,Min}} + \rho_{\text{ext,Max}} = 2\rho_{\text{ext,Av}}$, giving us

$$\langle \rho_{\text{ext,Max}} \rangle = 2\frac{C_i \alpha_s}{\pi^2} Q - 2\left(\frac{C_i \alpha_s}{\pi^2}\right)^2 A_{\text{Trans}}\, Q\,. \tag{21}$$

To appreciate the impact of the various terms, let us take $C_i \equiv C_A = 3$, $Q = 50\,\text{GeV}$, $\alpha_s(Q) = 0.13$, and $A_{\text{Trans}} = 2\pi/3$. Then we obtain

$$\langle \rho_{\text{ext,Av}} \rangle \simeq 2.0\,\text{GeV}, \tag{22a}$$

$$\langle \rho_{\text{ext,Min}} \rangle \simeq 0.3\,\text{GeV}, \tag{22b}$$

$$\langle \rho_{\text{ext,Max}} \rangle \simeq 3.6\,\text{GeV}. \tag{22c}$$

These numbers scale roughly linearly with Q. The crudeness of our approximations for the perturbative part of the event means that they are not be trusted to better than within a factor of two (worse in the case of $\langle \rho_{\text{ext,Min}} \rangle$). However the rough orders of magnitude are still instructive and highlight the advantage of the Min region.

The above analytic estimates can be verified by using more realistic events from a Monte Carlo generator at parton level. For the case of dijets from Herwig pp collisions at $\sqrt{s} = 10\,\text{TeV}$, with p_t of the partons in the hard 2→2 process required to be above $100\,\text{GeV}$ (consistent with $Q = 50\,\text{GeV}$) and the soft underlying event turned off, one obtains

$$\langle \rho_{\text{ext,Av}}^{\text{MC}} \rangle \simeq 2.1\,\text{GeV}, \tag{23a}$$

$$\langle \rho_{\text{ext,Min}}^{\text{MC}} \rangle \simeq 0.5\,\text{GeV}, \tag{23b}$$

$$\langle \rho_{\text{ext,Max}}^{\text{MC}} \rangle \simeq 3.8\,\text{GeV}. \tag{23c}$$

These numbers are very close to those from the simple model of purely perturbative underlying event described above for gluon jets (though Herwig has an admixture of quark jets here). We have verified that if we double the area over which the measurement is carried out, the $\langle \rho_{\text{ext,Min}}^{\text{MC}} \rangle$ result roughly doubles, as expected from eq. (20).

One comment concerning the above results is that in the pure soft UE case it was $\rho_{\text{ext,Av}}$ that was the least biased estimate of ρ. Here it is $\rho_{\text{ext,Min}}$ that is the least biased by hard perturbative radiation. If one restricts one's attention to $\rho_{\text{ext,Min}}$, then a further property of interest is that in the soft UE case, the bias is reduced by increasing the transverse region's area A_{Trans}, while for hard perturbative contamination increasing A_{Trans} increases the bias. This trade-off between the two issues is characteristic of the difficulty of accurately estimating ρ.

Area/median-based approach Let us suppose that we extract ρ based on jets contained in a region of area A_{tot}. Assuming the typical area for jets as introduced in eq. (13), then the typical number of jets N in the region (including ghost jets) should be given by

$$N \simeq \frac{A_{\text{tot}}}{c_J R^2}\,. \tag{24}$$

The exact value of N in each given event will depend on that event's detailed structure (and the exact set of ghosts), but eq. (24) should be adequate for our illustrative discussion here.

Of the N jets, we will assume that n_h are "hard" jets, of which n_b correspond to the final-state born particles ($n_b = 2$ for dijet events, $n_b = 0$ for Drell-Yan events) and n_p stem from perturbative radiation. It is convenient, albeit somewhat simplistic, to model n_p as being given by the number of emitted gluons[6]

$$\langle n_h \rangle = n_b + \langle n_p \rangle \simeq n_b + \frac{C_i}{\pi^2} \int_{A_{\text{tot}}} d\phi \, dy \int_{Q_0}^Q \frac{dp_t}{p_t} \alpha_s(p_t) = n_b + A_{\text{tot}} \frac{C_i}{\pi^2} \frac{1}{2b_0} \ln \frac{\alpha_s(Q_0)}{\alpha_s(Q)}, \qquad (25)$$

where we consider the number of perturbative emissions between some non-perturbative scale Q_0 and an upper limit Q related to the hard scale of the process (e.g. half the p_t of the hardest jet, as before) and we have used a 1-loop running approximation for $\alpha_s(p_t)$.

For the median estimator of ρ to be non-zero, at least half the jets should contain perturbative radiation, i.e. $n_p + n_b \geq N/2$. Since the number of primary emissions follows a Poisson distribution, we get the probability of non-zero ρ from the following sum

$$P = \sum_{n=N/2-n_b}^{N} \frac{\langle n_p \rangle^n e^{-\langle n_p \rangle}}{n!} \simeq \frac{\langle n_p \rangle^{(N/2-n_b)} e^{-\langle n_p \rangle}}{(N/2 - n_b)!}, \qquad (26)$$

where we have also made the approximation that the sum is dominated by its first term, on the grounds that $\langle n_p \rangle / (N/2 - n_b) \ll 1$. Given P, one can estimate ρ_{ext} by observing that the $(N/2)^{\text{th}}$ jet will be the softest of all the perturbative jets, and therefore have $p_t \sim Q_0$, giving

$$\langle \rho_{\text{ext}} \rangle \simeq \frac{Q_0}{\langle A_{\text{jet}} \rangle} P \simeq \frac{Q_0}{c_J R^2} P. \qquad (27)$$

This is plotted with thick lines in fig. 4 (left) as a function of A_{tot} for $R = 0.6$, using a 1-loop 5-flavour coupling with $\Lambda = 0.1 \, \text{GeV}$ [7] and with $C_i = C_A = 3$, $Q = 50 \, \text{GeV}$, $Q_0 = 1 \, \text{GeV}$ and two values for n_b, 0 and 2.

To estimate the uncertainties of our analytic formula introduced by approximations (26) and (27) we also plot with thin lines in fig. 4 (left) the result of numerical studies of the same simple set of perturbative emissions (equivalent to the full sum over n). As before, we associate each parton (i.e. gluon in our case) with one jet and assume that this jet has area $c_J R^2$. In the case $n_b = 2$, we make sure that the two Born particles are always present in the region where the underlying event is measured. One sees in fig. 4 that the exact numerical results (equivalent to taking the full sum in eq. (26)) are moderately higher than the corresponding analytic approximations, as would be expected. In both cases, however, the contribution to ρ_{ext} is negligible except for small values of A_{tot} with $n_b = 2$.

The approximation that we have used for the distribution of the number of jets is rather simplistic. To estimate the order of magnitude of the uncertainties that are present in our toy model of the perturbatively induced "UE", notably those associated with secondary radiation from the partons, we have also applied the area/median method to realistic dijet events from Herwig with soft UE turned off, taken at parton level. To make the comparison, we required

---

[6]Recall that our model includes only primary emissions from the incoming partons — we neglect cases where two of these emissions end up in the same jet (which would decrease n_p), and also the fragmentation contribution from the Born and radiated particles (mostly contained within the respective jets, but which would increase n_p in the cases where the fragmentation is outside those jets).

[7]This mimics a 2-loop coupling with $\alpha_s(M_Z) = 0.120$ to within a couple of percent over a broad range of scales.
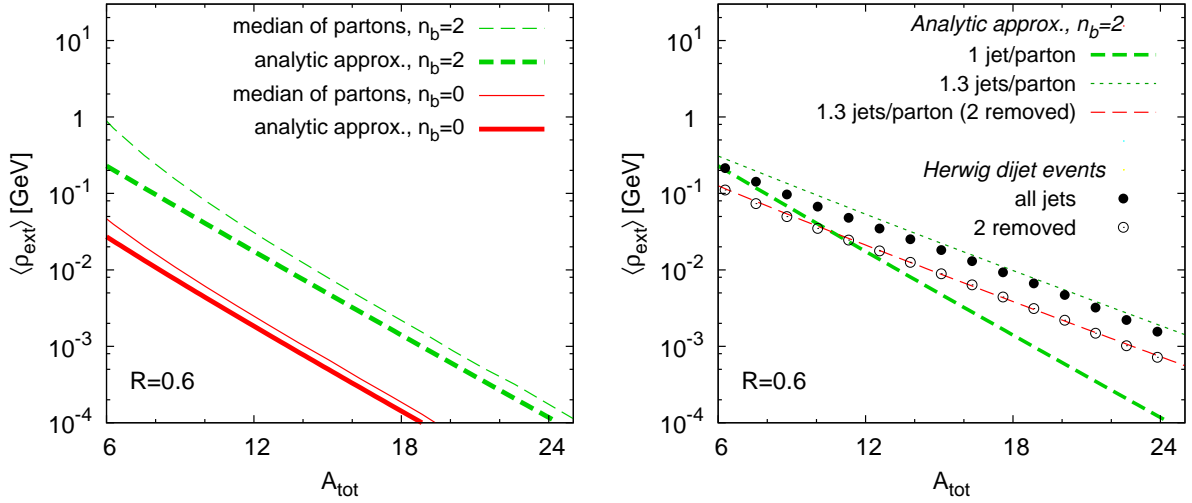
Figure 4: Left: pure perturbative underlying event extraction via the area/median method, as a function of the total area, for cases with zero and two Born particles. The curves correspond to the simple model of perturbative emissions described in the text. Thick lines show the analytic estimate of ρ_{ext} based on the RHS of eq. (26) and thin lines correspond to the use of the median on our model for the ensemble of parton emissions (see text for further details). Right: comparison of the analytical prediction for the case with 2 Born particles with results from parton-level (no UE) events generated with Herwig, and clustered with C/A algorithm (points). Two analytical results are shown: one with $\langle n_p \rangle$ as given in eq. (25) (1 jet/parton) and another in which $\langle n_p \rangle \to 1.3 \langle n_p \rangle + 0.3 n_b$, i.e. each parton generates an average of 1.3 jets. The figure also shows a curve (analytical) and points (Herwig) that illustrate the impact of removing the two hardest jets from the list used in determining the median.

that the partons in the hard $2 \to 2$ process have $p_t > 100$ GeV, as before (and consistent with our choice $Q = 50$ GeV). The result is shown in fig. 4 (right) with black and open circles for the cases with and without removal of the two hardest jets from the ensemble used for the median. The A_{tot}-dependence of the Herwig results differs somewhat from the expectations in eqs. (26) and (27). We attribute this to the fact that each parton (be it a Born parton or a primary emission) can itself radiate extra gluons, some of which will lead to additional jets. A simple way of accounting for this is to replace $\langle n_p \rangle$ in eq. (26) with $(1 + X) \langle n_p \rangle + X n_b$, where X is the number of extra jets obtained per parton.[8] Using a modest $X = 0.3$ (1.3 jets per parton) brings the analytical result into accord with the Herwig results.

In terms of the practical impact of perturbative emission on the extracted ρ values, one sees that for $A_{\text{tot}} \geq 4\pi$ (the minimal value that we shall use in section 4) it remains a small effect, though the curves also highlight that at small total area A_{tot}, the area/median method can start to become sensitive to perturbative radiation, especially when Born partons are present.

A final point to comment on is the relation between these results and the discussion in [13], where it was argued that ρ_{ext} would be non-zero starting only at $\mathcal{O}(\alpha_s^{N/2-n_b})$. Examining eq. (26), one sees that this statement stems from the fact that $\langle n_p \rangle \sim \alpha_s$ (in a fixed-coupling

---

[8]A more sophisticated approach would be to calculate X using soft-gluon resummation, which would lead to a single-logarithmic dependence of X on the parton transverse momentum.

13

approximation). However, each power of α_s is compensated in part by a power of $A \ln Q/Q_0$ (which is large), and ultimately the small value of P in eq. (26) (and hence $\langle \rho_{\text{ext}} \rangle$) cannot be solely attributed to an α_s power-counting argument, but rather involves a more subtle interplay of all the factors in eq. (27).

3.3 Two-component events

Realistic events are neither purely perturbative nor consist of pure soft noise. It is instructive to examine what happens if one considers events that have both components together.

Traditional approach. The transverse-momentum density extracted from the average of the two transverse regions is straightforward to calculate in the two component model: it is just given by the sum of the soft and perturbative components,

$$\langle \rho_{\text{ext,Av}} \rangle = \rho + \frac{C_i \alpha_s}{\pi^2} Q \,. \tag{28}$$

The results for the Min and Max regions are more complex: in the pure soft-component case, it was the soft radiation that determined which of the two regions was Min/Max; analogously, in the perturbative case, it was the perturbative radiation that determined this. When both can be present one has to consider which of the two fixes the Min/Max regions. It is useful to define $\mathcal{P}$ to be the fraction of events in which the amount of perturbative radiation in each of the two transverse regions is smaller than the size of soft fluctuations of those regions, $p_{t,L}, p_{t,R} \ll \sigma \sqrt{A_{\text{Trans}}}$. In this set of events, it is the soft component that defines which region is Min/Max, and the bias in the extraction of ρ is just the soft bias, eqs. (9), with no perturbative bias. In the remaining events, in which one or more perturbative emissions are much harder than $\sigma \sqrt{A_{\text{Trans}}}$, it is those perturbative emissions that will determine which of the two regions is Min/Max. For these events, there will be no bias in the contribution from the soft component. This implies that the average bias in the Min/Max regions for the soft component will be $\mp \mathcal{P} \cdot \sigma / \sqrt{\pi A_{\text{Trans}}}$.

As concerns the perturbative contamination, the average results in eqs. (20) and (21) are already dominated by the set of events in which there is at least one hard emission, so these contributions remain unchanged in the two-component case. The final result for the Min/Max regions is therefore

$$\langle \rho_{\text{ext,Min}} \rangle \simeq \rho - \frac{\sigma \mathcal{P}}{\sqrt{\pi A_{\text{Trans}}}} + 2 \left(\frac{C_i \alpha_s}{\pi^2} \right)^2 A_{\text{Trans}} \, Q \,, \tag{29a}$$

$$\langle \rho_{\text{ext,Max}} \rangle \simeq \rho + \frac{\sigma \mathcal{P}}{\sqrt{\pi A_{\text{Trans}}}} + 2 \frac{C_i \alpha_s}{\pi^2} Q - 2 \left(\frac{C_i \alpha_s}{\pi^2} \right)^2 A_{\text{Trans}} \, Q \,, \tag{29b}$$

where $\mathcal{P}$ is given by

$$\mathcal{P} \simeq \exp \left(-2 A_{\text{Trans}} \frac{C_i}{\pi^2} \frac{1}{2 b_0} \ln \frac{\alpha_s(\max(Q_0, \sigma \sqrt{A_{\text{Trans}}}))}{\alpha_s(Q)} \right) \,. \tag{30}$$

Note that the choice of lower scale in the logarithm, $\sigma \sqrt{A_{\text{Trans}}}$, or Q_0 if that is larger, is only controlled to within a factor of $\mathcal{O}(1)$, just as is the choice of Q for the upper limit.

Area/median-based approach. The combination of low-p_t and perturbative components is non-trivial also in the case of the area/median-based approach. To treat it analytically, it will be convenient to work at R values that are sufficiently large that the distribution of $p_{t,\text{jet}}/A_{\text{jet}}$ for the many jets can be considered approximately Gaussian — i.e. we will work away from the turn-on region $\nu A_{\text{jet}} = \ln 2$ that was discussed in section 3.1.

Of the N jets that are used in determining the median, some will contain hard perturbative radiation with transverse momentum significantly above the scale of the fluctuations of the UE. Assuming that there are on average $\langle n_h \rangle$ hard partons, and that the probability distribution of hard partons in a jet (or tile) is given by a Poisson distribution with mean $\langle n_h \rangle /N$, then the average number of jets not contaminated by the hard partons will be given by $N e^{-\langle n_h \rangle /N} \simeq N - \langle n_h \rangle + \langle n_h \rangle^2 /(2N)$. These uncontaminated jets will have a distribution of values of $\rho_{\text{jet}} = p_{t,\text{jet}}/A_{\text{jet}}$ that is governed just by the soft component and is roughly Gaussian

$$\frac{dn^{(\text{soft})}}{d\rho_{\text{jet}}} = \frac{1}{\sigma}\sqrt{\frac{\langle A_{\text{jet}}\rangle}{2\pi}} \exp\left(-\frac{\langle A_{\text{jet}}\rangle}{2\sigma^2}(\rho_{\text{jet}} - \rho)^2\right) \cdot N \exp\left(-\frac{\langle n_h \rangle}{N}\right). \tag{31}$$

Assuming $\langle n_h \rangle < N \ln 2$, the median procedure implies finding ρ_{ext} such that $N/2$ of the $N e^{-\langle n_h \rangle /N}$ Gaussian-distribution jets have $\rho_{\text{jet}} < \rho_{\text{ext}}$, i.e. one must determine the value of ρ_{ext} such that

$$\int_{-\infty}^{\rho_{\text{ext}}} d\rho_{\text{jet}} \frac{dn^{(\text{soft})}}{d\rho_{\text{jet}}} = \frac{N}{2}, \tag{32}$$

(the unphysical negative lower limit of the integral, an artefact of the Gaussian approximation, doesn't perturb the argument). In the small $\langle n_h \rangle /N$ limit, this is easily solved and gives

$$\langle \rho_{\text{ext}}\rangle \simeq \langle \rho_{\text{ext}}^{(\text{soft})}\rangle + \sigma\sqrt{\frac{\pi}{2\langle A_{\text{jet}}\rangle}} \left(\frac{\langle n_h \rangle}{N} + \frac{\langle n_h \rangle^2}{2N^2} + \mathcal{O}\left(\frac{\langle n_h \rangle^3}{N^3}\right)\right), \tag{33}$$

where $\langle \rho_{\text{ext}}^{(\text{soft})}\rangle$ is the result obtained in the pure soft case, eq. (11) of section 3.1.[9] One can then use eqs. (13) and (24) to express $\langle A_{\text{jet}}\rangle$ and N in terms of R and A_{tot}. Keeping only the first two terms in the R expansion gives

$$\langle \rho_{\text{ext}}\rangle \simeq \langle \rho_{\text{ext}}^{(\text{soft})}\rangle + \sqrt{\frac{\pi c_J}{2}}\sigma R \left(\frac{\langle n_h \rangle}{A_{\text{tot}}} + c_J R^2 \frac{\langle n_h \rangle^2}{2A_{\text{tot}}^2}\right). \tag{34}$$

Features to note here are that the discrepancy is proportional to σR, with the next correction going as σR^3. Eq. (25) provides a result for the average number of hard emissions $\langle n_h \rangle$, and Q_0 there should be replaced with $\max(Q_0, \mathcal{O}\left(\sigma\sqrt{A_{\text{jet}}}\right))$, or equivalently $\max(Q_0, \sqrt{c_J}\sigma R)$, because perturbative emissions whose transverse momentum is much smaller than the scale of fluctuations of the underlying event will not bias the median. This then gives us

$$\frac{\langle n_h \rangle}{A_{\text{tot}}} \simeq \frac{n_b}{A_{\text{tot}}} + \frac{C_i}{\pi^2}\frac{L}{2b_0}, \qquad L \equiv \ln\frac{\alpha_s(\max(Q_0, \sqrt{c_J}\sigma R))}{\alpha_s(Q)}. \tag{35}$$

We see that the Born particles contribute when A_{tot} is not very large. Perturbative emissions always contribute, essentially because the number of emissions scales with the total area.

---

[9]The additivity of soft and hard results is an approximation, justified only when $\langle \rho_{\text{ext}}^{(\text{soft})}\rangle$ is close to ρ. An additional point is that when plotting soft+hard results for $\langle \rho_{\text{ext}}\rangle$, we will eliminate the Θ-function in eq. (11) and use the prescription that $\langle \rho_{\text{ext}}\rangle$ is well-defined only when the sum of soft and hard contributions gives a positive result.

Substituting physically reasonable numbers into eq. (35), i.e. which corresponds to setting the logarithm L equal to 1, gives

$$\frac{\langle n_h \rangle}{A_{\text{tot}}} \simeq \frac{n_b}{A_{\text{tot}}} + 0.25 \frac{C_i}{C_A} \, , \qquad (36)$$

which then gives the following numerical result for the bias,

$$\langle \rho_{\text{ext}} \rangle \simeq \langle \rho_{\text{ext}}^{(\text{soft})} \rangle + 0.45 \, \sigma R \cdot \left(\frac{C_i}{C_A} + 4.0 \frac{n_b}{A_{\text{tot}}} \right) + 0.11 \, \sigma R^3 \cdot \left(\frac{C_i}{C_A} + 4.0 \frac{n_b}{A_{\text{tot}}} \right)^2 \, . \qquad (37)$$

Ignoring the n_b/A_{tot} terms, for $R = 0.6$ in gluon-initiated processes ($C_i = C_A$), the bias introduced in ρ is about 0.29σ. For values of $\sigma = 0.5\rho - 0.7\rho$, as we will obtain from the Monte Carlo study in section 4, the positive bias due to these perturbative effects is in the same $15 - 20\%$ ballpark as the negative bias due to finite-particle density that was discussed in section 3.1 for pure soft events. There is an expectation that these two sources of bias should combine linearly, at least when R is sufficiently far above the critical turn-on point that the Gaussian approximation used above is valid. Since they have different R-dependences, $\sim +R$ and $\sim -1/R^2$ respectively, there exists only a limited range of R in which they compensate each other. In this respect the numbers given above tend to confirm the choice $R \sim 0.5 - 0.6$ that had originally been recommended based on Monte Carlo studies in [13].[10]

Fig. 5 summarises the study of the two component model by showing the analytic results from eqs. (34) and (36) together with the numerical results from the median tile-based and median jet-based approaches. Two sets of results are presented, corresponding to the events with and without Born particles. We see the very good agreement between the analytic and the numerical tile-based approaches and the median jet-based result (except in the threshold region, as in section 3.1). To achieve this it was essential to include the term $\sim R^3$ in eq. (34).

Fig. 5 suggests that the region of R with the least bias for the determination of ρ is $R = 0.4 - 0.6$. If one requires that the biases in eqs. (15) and (34) cancel each other, then one finds that R should be chosen proportional to $\nu^{1/6}$ or equivalently proportional to $(\sigma/\rho)^{1/3} \sim R_{\text{crit}}^{1/3}$. Ignoring the n_b/A_{tot} component and the R^3 term in the equations of this section, one finds

$$R_{\text{zero-bias}} \simeq \sqrt{\frac{\pi}{2^{\frac{1}{3}} c_J}} \left(\frac{\sigma}{\rho} \frac{b_0}{C_i L} \right)^{\frac{1}{3}} \simeq 0.65 \left(\frac{\sigma}{\rho} \frac{C_A}{C_i} \right)^{\frac{1}{3}} \simeq 0.87 \, R_{\text{crit}}^{\frac{1}{3}} \left(\frac{C_A}{C_i} \right)^{\frac{1}{3}} \, , \qquad (39)$$

where the numerical values have been obtained setting $L = 1$ in eq. (35). The result for $R_{\text{zero-bias}}$ can be seen to be consistent with fig. 6, which shows the analytical approximation for ρ_{ext} as a function of R for a broad range of particle densities ν. For a variation in the

---

[10] An alternative method of extraction of ρ comes to mind at this point, fitting a formula motivated by the results of this section and appendix A:

$$\rho_{\text{ext}} = \rho \left(\frac{R^2 - R_{\text{crit}}^2}{R^2 + R_{\text{crit}}^2 \left[\frac{n}{(n+1)\ln 2} - 1 \right]} \right)^{\frac{1}{n}} (1 + cR) \, , \qquad (38)$$

with the fit parameters ρ, R_{crit}, n, c. One might choose to forgo n, or try a finite number of choices, e.g. $n = 1, 2$. We have found that such a procedure eliminates a substantial part of the biases in the extraction of ρ for some events, but in others statistical fluctuations lead to poor fits, with results for ρ that are highly skewed towards low values. For this reason, we do not adopt this approach for the current study.
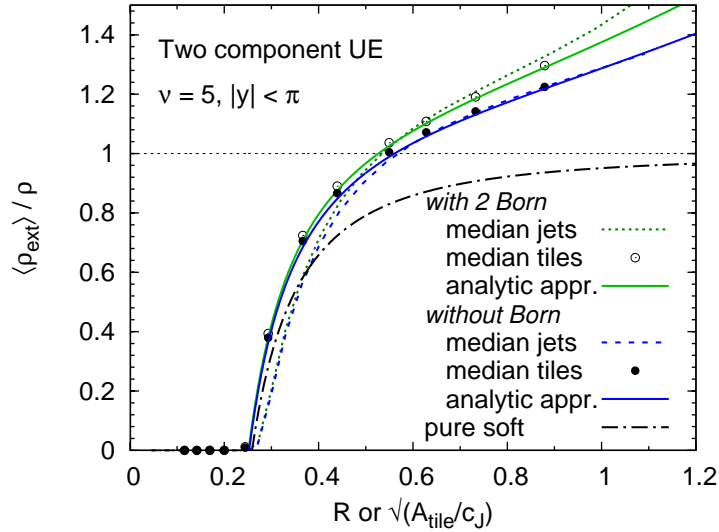
Figure 5: Underlying event from the two component toy model analysed in the range of rapidity $|y| < \pi$ and the full span of the azimuthal angle. The results for the average ρ_{ext} (normalised to ρ), extracted in tile and jet median-based approach (see text) are shown for two sets of events. The first set contains soft and perturbative particles while in the second set, two Born particles with $p_{t,\text{hard}} = 100\,\text{GeV}$ are present in addition. For reference, the analytic curve from fig. 3 for the pure soft case is also shown. The mean number of soft particles per unit area, ν, was 5 and their average transverse momentum $\mu = 0.4\,\text{GeV}$, which corresponds to $\rho = 2\,\text{GeV}$. The perturbative emissions are distributed between scales $Q_0 = 1\,\text{GeV}$ and $Q = 50\,\text{GeV}$.

particle density (and ρ) by a factor of 50 (σ/ρ by a factor of 7), the zero-bias R value changes only moderately and in close accord with the expectations of eq. (39). Fig. 6 also illustrates that a fixed $R \sim 0.6$ leads to a ρ_{ext} value to within 20% of the correct result for a whole range of ν, with the relative impact of the biases steadily decreasing as the particle density is increased.

3.4   Fluctuations in the estimation of ρ

In the simple model studies discussed here, the same ρ value has been used to generate all events. Nevertheless, extracted values of ρ vary from one event to another. This is because any method of measuring UE can use only a limited part of an event (restricted A_{tot}, A_{Trans}, according to y and ϕ cuts) and works with a finite number of objects (particles, jets). The magnitude of the observed event-to-event fluctuations in this case is an important characteristic of a method, because it sets a lower limit on the uncertainty of the event-to-event ρ measurement. These intrinsic fluctuations also affect the measurement of the true fluctuations and correlations of realistic UE models. Therefore, in order to measure properties of the underlying event that can then be used for efficient subtraction or tuning of simulation programs, one is interested in reducing the fluctuations that come with the method itself.

For the soft underlying event from section 3.1, ignoring the small systematic biases that we found there in the determination of ρ, one can show that the standard deviations S_d of
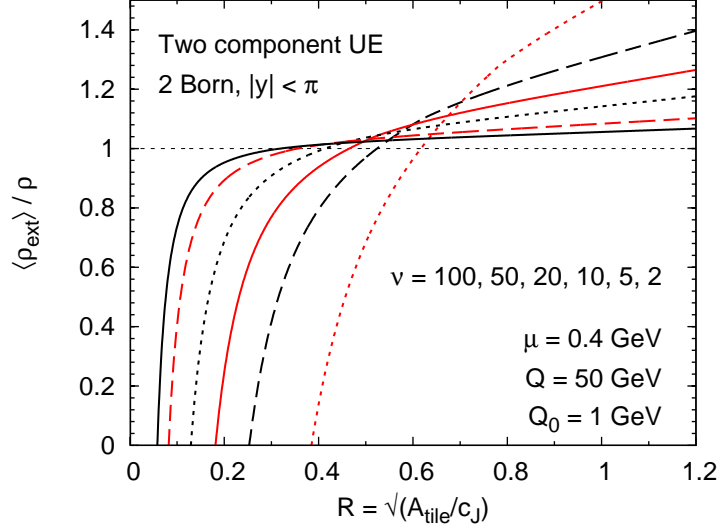
Figure 6: The analytical approximation for ρ_{ext} as a function of R, eqs. (11,34), shown for a range of soft-particle densities (ν values), for the gluonic case, $C_i = C_A$. The curves from left to right (at small R values) correspond to $\nu = 100, 50, 20, 10, 5, 2$ (or equivalently ρ from $40\,\text{GeV}$ to $0.8\,\text{GeV}$).

the extracted ρ values read

$$\frac{S_{d,\text{med}}^{(\text{soft})}}{\rho} = \sqrt{\frac{\pi}{\nu A_{\text{tot}}}} = \frac{\sigma}{\rho}\sqrt{\frac{\pi}{2 A_{\text{tot}}}}\,, \tag{40a}$$

$$\frac{S_{d,\text{Av}}^{(\text{soft})}}{\rho} = \sqrt{\frac{1}{\nu A_{\text{Trans}}}} = \frac{\sigma}{\rho}\frac{1}{\sqrt{2 A_{\text{Trans}}}}\,, \tag{40b}$$

$$\frac{S_{d,\text{Min}}^{(\text{soft})}}{\rho} = \frac{S_{d,\text{Max}}^{(\text{soft})}}{\rho} = \sqrt{\frac{2}{\nu A_{\text{Trans}}}\left(1 - \frac{1}{\pi}\right)} = \frac{\sigma}{\rho}\sqrt{\frac{\pi - 1}{\pi A_{\text{Trans}}}}\,. \tag{40c}$$

The result for the median case is derived in appendix B.1. The formula for the Min/Max regions was obtained in the Gaussian approximation, valid in the limit $\nu A_{\text{Trans}} \gg 1$. Substituting realistic values for the areas, i.e. $A_{\text{Trans}} = 2\pi/3$, $A_{\text{tot}} = 4\pi$, and the density $\nu = 5$, one arrives at the following numerical estimates

$$S_{d,\text{med}}^{(\text{soft})}/\rho = 0.22, \qquad S_{d,\text{Av}}^{(\text{soft})}/\rho = 0.31, \qquad S_{d,\text{Min/Max}}^{(\text{soft})}/\rho = 0.36. \tag{41}$$

The lower expected fluctuations for the area/median based approach are a consequence of the larger area used in the UE determination. The use of a larger area is possible in the first place because the method's dynamical separation of UE limits the need to cut away regions from the $y - \phi$ plane to reduce contamination from hard jets.

In fig. 7, we show histograms of ρ extracted in the traditional and area/median approaches for the case of purely soft underlying event from section 3.1 (left) and for the combined model described in section 3.3 with the soft and perturbative components and 2 Born particles (right). The corresponding standard deviations of the ρ_{ext} values are given in table 1. For the case of the purely soft UE, we see that these standard deviations follow the pattern from

18

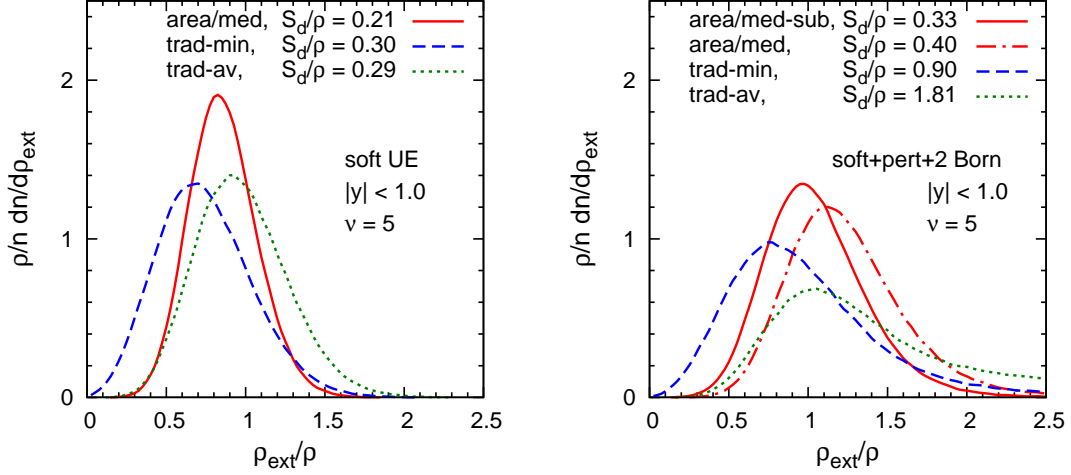Figure 7: Distributions of ρ_{ext} from traditional and area/median based approach for the case of pure soft underlying event from section 3.1, (left) and for the combined model from section 3.3 with the soft and perturbative components and 2 Born particles (right). The average density of soft particles is $\nu = 5$ and their average transverse momentum is $\mu = 0.4\,\text{GeV}$, which corresponds to $\rho = 2\,\text{GeV}$. The perturbative emissions are distributed between scales $Q_0 = 1\,\text{GeV}$ and $Q = 50\,\text{GeV}$. To guarantee that the two hardest jets are associated with the Born particles, their transverse momenta were set to $p_t = 400\,\text{GeV}$. The jet clustering was performed with the Cambridge/Aachen algorithm with R=0.6.

	$S_{d,\text{ext}}/\rho$	
	pure soft	soft + PT + 2 born
median/area, all jets	0.21	0.40
median/area, all-2 jets	–	0.33
traditional, TransAv	0.30	1.81
traditional, TransMin	0.29	0.90

Table 1: Results for $S_{d,\text{ext}}/\rho$ for various methods of extracting ρ in the toy model where the input standard deviation of ρ is $S_d = 0$. These values provide an indication of the lower bound of S_d values that could be observed in real events.

eq. (41), with the area/median approach performing best. Note however that finite-density (ν) effects do tend to slightly reduce the standard deviation results as compared to expectations, especially in the TransMin case.

Adding perturbative and Born particles increases fluctuations in all cases. From the shapes of the curves in fig. 7 (right), it seems that the area/median and TransMin results are both degraded by similar amounts, while the TransAv result suffers significantly more, with a long tail to large ρ_{ext} values. The strong degradation of the TransAv result is an expected consequence of its sensitivity to perturbative radiation,

$$S_{d,\text{Av}}^{(\text{hard})} \simeq \sqrt{\frac{C_i \alpha_s}{4 A_{\text{Trans}} \pi^2}} Q \left(1 + \mathcal{O}\left(\sqrt{\alpha_s A_{\text{Trans}}} \right) \right), \tag{42}$$

where one observes the dependence on $\sqrt{\alpha_s / A_{\text{Trans}}}$ (to be compared to the bias on the mean which goes as α_s).

However, if one examines the results for S_d in table 1, one sees that the TransMin standard deviation is also significantly increased by perturbative radiation. The toy-model expectation is

$$S_{d,\text{Min}}^{(\text{hard})} \simeq \frac{C_i \alpha_s}{\pi^2 \sqrt{2}} Q \left(1 + \mathcal{O}\left(\alpha_s^2 A_{\text{Trans}}^2 \right) \right), \tag{43}$$

and for the particular parameters used in table 1, the result for S_d turns out to be as large as ρ itself. Compared to the $\mathcal{O}\left(\alpha_s^2 A \right)$ suppression for the bias to the $\langle \rho_{\text{ext,Min}} \rangle$ result, here the perturbative radiation bias is much stronger, $\mathcal{O}\left(\alpha_s \right)$. The physical explanation is simple: while it is relatively rare for perturbative radiation to affect the TransMin region (hence the acceptable peak-region of the ρ_{ext} distribution), when it does, the effect on ρ_{ext} is large, contribution significantly to the final result for S_d.[11]

The area/median approach is much more robust in this respect, because the hard emissions' contribution to the standard deviation does not have significant enhancements compared to the average bias on ρ:

$$S_{d,\text{med}}^{(\text{hard})} \simeq 1.79 \frac{\sigma R}{\sqrt{A_{\text{tot}}}} \left(\frac{C_i}{C_A} + 4.0 \frac{n_b}{A_{\text{tot}}} \right)^{\frac{1}{2}} + \cdots, \tag{44}$$

as derived in appendix B.2. In particular, the larger numerical coefficient compared to the $\mathcal{O}\left(R \right)$ term of eq. (37) is compensated by the factor of $1/\sqrt{A_{\text{tot}}}$. This good behaviour is visible in fig. 7, and also in the values of table 1, which are roughly consistent with the above analytical estimate (maybe 20% higher). They also highlight the further improvement to be had with the hard-jet removal procedure discussed at the end of section 2.2 ("all-2 jets" result), which benefits not just S_d, but also the peak position and height in fig. 7.

Overall, the results of this section suggest that for any measurement of fluctuations of the UE, it will be preferable to use the area/median method, with hard-jet removal able to provide some extra benefit.

---

[11]The hard component also has an impact on the soft fluctuations, in analogy with the effect discussed for $\langle \rho_{\text{ext,Min}} \rangle$ in section 3.3. We ignore this complication here since the fluctuations of the hard component in any case dominate over those of the soft component.
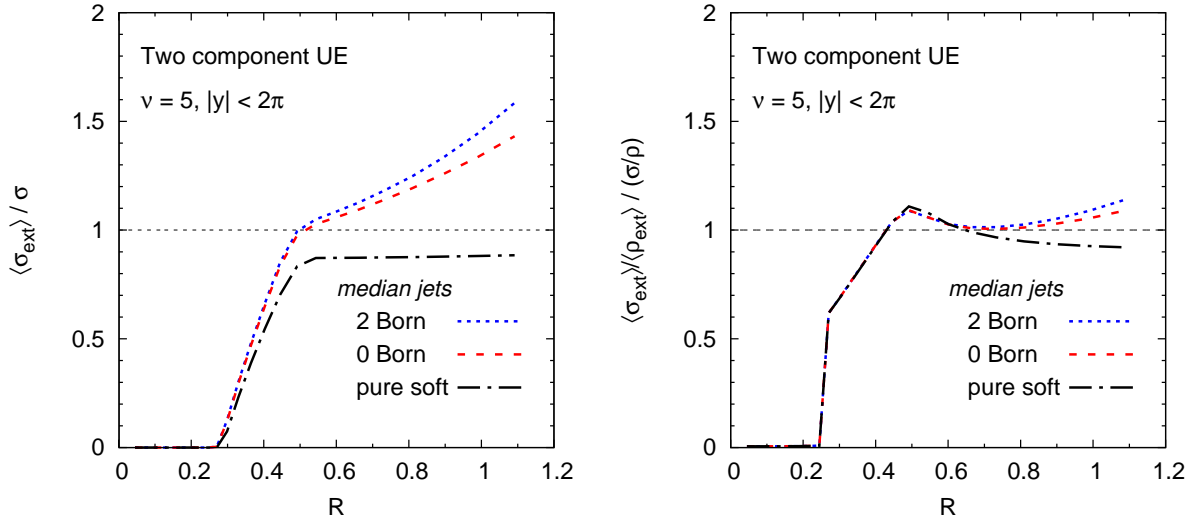
Figure 8: Left: extracted value of σ with the area/median approach in the pure soft and the two component toy models, shown as a function of R and normalised to the correct value for the soft component $\sigma_{\mathrm{soft}} = \rho_{\mathrm{soft}}\sqrt{2/\nu}$. Right: the same set of results, but showing $\langle\sigma_{\mathrm{ext}}\rangle / \langle\rho_{\mathrm{ext}}\rangle$ normalised to the expected σ/ρ for the soft component. The toy model variant is that corresponding to eq. (4).

3.5 Extraction of σ

The measurement of intra-event fluctuations, σ, has only been discussed so far in the context of the area/median approach.[12] Though we shall not go into full analytical detail, it is easy to convince oneself that many of the considerations that arise in the extraction of ρ apply also when determining σ. In particular, for pure soft events one underestimates σ when R is too small and the presence of perturbative radiation will bias the extracted σ at larger R.

One point to be aware of is that our method for extracting it, cf. section 2.2, has the characteristic that $\sigma/\sqrt{A_{\mathrm{jet}}}$ never exceeds ρ. This translates into an R-dependent upper bound on $\sigma_{\mathrm{ext}}/\rho_{\mathrm{ext}}$,

$$\frac{\sigma_{\mathrm{ext}}}{\rho_{\mathrm{ext}}} \lesssim \sqrt{c_J}\, R \simeq 1.43R\,. \tag{45}$$

Fig. 8 (left) shows the average σ_{ext} as a function of R in the toy model ($\nu = 5$, with and without the perturbative radiation and the Born jets), normalised to the correct result for the soft component. In the case of just the soft component, one sees a threshold region followed by a slow approach towards the correct value, much as for ρ_{ext} even if the detailed shapes differ (in part owing to the one-sided determination of σ, which causes the residual bias to be proportional to $1/R$, rather than $1/R^2$ for the bias to ρ). With the inclusion of the perturbative component there is an additional bias, which grows towards large R, again much as happens for ρ_{ext}.

To compare the biases on σ_{ext} and ρ_{ext}, it is convenient to examine the ratio $\langle\sigma_{\mathrm{ext}}\rangle / \langle\rho_{\mathrm{ext}}\rangle$, fig. 8 (right), normalised to the soft-component result for σ/ρ. The impact of the bound on

---

[12]In the traditional approach, one might envision calculating the perturbative contributions in eqs. (28,29) to NLO with a program such as NLOJet++ [33] and then fitting for $\langle\rho\rangle$ and $\langle\sigma\rangle$ in those equations after averaging over many events. It seems that it would be difficult, however, to extract event-by-event estimates of σ.

σ/ρ, eq. (45), is clearly visible up to $R \simeq 0.5$ (consistent, roughly, with $\sigma/\rho \simeq 0.63$). Beyond this point, over a reasonable range of R, $\langle \sigma_{\text{ext}} \rangle / \langle \rho_{\text{ext}} \rangle$ remains compatible with the true value to within roughly $10 - 20\%$, and, as expected, deviations are larger in the presence of hard radiation. Overall, in the region of R that is suitable for extracting ρ, figure 8 suggests that the extraction of σ should also be quite acceptable.

3.6   Summary of main results

Table 2 summarises the main results of this section for the biases, $\delta\rho$, and the event-to-event fluctuations, S_d, that occur within the toy model in extracting ρ, the transverse momentum flow per unit area, with each of three UE estimation methods: the traditional approach in its TransAv and TransMin variants, whose main parameter is the area A_{Trans} of each of the two transverse regions; and the area/median approach whose parameters are the total area under consideration, A_{tot} and the jet radius R. Results are given both for the biases and fluctuations intrinsic to the soft component and for the additional biases that arise due to the presence of hard radiation in the event, expressed as

$$\rho_{\text{ext}} = \rho + \delta\rho^{(\text{soft})} + \delta\rho^{(\text{hard})} \pm S_d^{(\text{soft})} \pm S_d^{(\text{hard})} \, . \tag{46}$$

The analytical formulae help illustrate the dependence on the parameters of the measurement methods (the area of the regions used, the jet radius) and the physical scales present in the events (hard scale Q, and the values for ρ and for the level of intra-event UE fluctuations, σ). One sees, for example, how the hard biases and event-to-event fluctuations in the traditional approach are always proportional to Q and to some power of α_s, whereas in the area/median approach they are proportional to σ (with modest coefficients and weak additional $\ln \ln Q$ dependence contained in the parameter L).

Table 2 also gives numerical results for the biases and S_d values, based on the default set of measurement parameter choices and physical scales that have been used throughout this section. This helps illustrate the expected orders of magnitude of different effects under realistic conditions. The large results for $S_d^{(\text{hard})}$ in the traditional method (i.e. unreliable event-by-event extraction of ρ), together with the proportionality to Q of the traditional method's biases and fluctuations and the fact that it offers no easy way of determining σ, lead us to prefer the area/median method for the Monte Carlo UE measurement studies that we will perform below.

Other results of this section that are not summarised in table 2 include: R_{crit}, eq. (14) the R-value below which the area/median approach gives $\rho_{\text{ext}} = 0$; the upper bound on $\sigma_{\text{ext}}/\rho_{\text{ext}}$ as a function of R in the median-area method, eq. (45); and $R_{\text{zero-bias}}$, eq. (39), the R value for which the soft and hard biases cancel out in the area/median approach. Finally, fig. 5 shows the characteristic shape of the R-dependence for ρ_{ext} in the median/area method, while fig. 6 helps illustrate how $R = 0.6$ is a reasonable default choice for a wide range of UE conditions.

4   Illustration with Monte Carlo events

Given the area/median method to determine ρ and σ on an event-by-event basis, let us now explore what kinds of observables we might construct from them. The choices that we shall make are motivated by considerations of how the UE affects jet measurements at hadron colliders. For example the UE leads to an average shift in jet energy; it's important that

quantity	method	result	numerical value	equation(s)
$\dfrac{\delta\rho^{(\text{soft})}}{\rho}$	TransAv	0	0	(9a)
	TransMin	$-\dfrac{\sigma}{\rho}\dfrac{\mathcal{P}}{\sqrt{\pi A_{\text{Trans}}}}$	-0.09	(9b), (29a)
	Area/Med*	$-\dfrac{\sigma^2}{\rho^2}\dfrac{1}{4c_J R^2}$	-0.14	(15)
$\dfrac{\delta\rho^{(\text{hard})}}{\rho}$	TransAv	$\dfrac{C_i\alpha_s}{\pi^2}\dfrac{Q}{\rho}$	0.99	(19), (28)
	TransMin	$2\left(\dfrac{C_i\alpha_s}{\pi^2}\right)^2\dfrac{A_{\text{Trans}}Q}{\rho}$	0.16	(20), (29a)
	Area/Med	$\dfrac{\sigma R}{\rho}\sqrt{\dfrac{\pi c_J}{2}}\left(\dfrac{n_b}{A_{\text{tot}}}+\dfrac{C_i}{\pi^2}\dfrac{L}{2b_0}\right)$	0.17	(34)–(37)
$\dfrac{S_d^{(\text{soft})}}{\rho}$	TransAv	$\dfrac{\sigma}{\rho}\sqrt{\dfrac{1}{2A_{\text{Trans}}}}$	0.31	(40b)
	TransMin	$\dfrac{\sigma}{\rho}\sqrt{\dfrac{\pi-1}{\pi A_{\text{Trans}}}}$	0.36	(40c)
	Area/Med	$\dfrac{\sigma}{\rho}\sqrt{\dfrac{\pi}{2A_{\text{tot}}}}$	0.22	(40a)
$\dfrac{S_d^{(\text{hard})}}{\rho}$	TransAv	$\sqrt{\dfrac{C_i\alpha_s}{4A_{\text{Trans}}\pi^2}}\dfrac{Q}{\rho}$	1.72	(42)
	TransMin	$\dfrac{C_i\alpha_s}{\pi^2\sqrt{2}}\dfrac{Q}{\rho}$	0.70	(43)
	Area/Med	$\dfrac{\sigma R}{\rho}\sqrt{\dfrac{2\pi c_J}{A_{\text{tot}}}}\left(\dfrac{n_b}{A_{\text{tot}}}+\dfrac{C_i}{\pi^2}\dfrac{L}{2b_0}\right)^{\frac{1}{2}}$	0.19	(44), (72)

Table 2: Summary of main biases and sources of fluctuations for different UE extraction methods; Q is the hard scale, Q_0 the IR cut-off on perturbative emissions, $L = \ln\left[\alpha_s(\max(Q_0,\sqrt{c_J\sigma R}))/\alpha_s(Q)\right]$, $\mathcal{P} \simeq \exp\left(-2A_{\text{Trans}}\frac{C_i}{\pi^2}\frac{1}{2b_0}\ln\left[\alpha_s(\max(Q_0,\sigma\sqrt{A_{\text{Trans}}}))/\alpha_s(Q)\right]\right)$, $c_J = 2.04$, C_i is the colour factor of the incoming partons. Numerical values are given for $\rho = 2\,\text{GeV}$, $\sigma/\rho = 0.63$, $Q = 50\,\text{GeV}$, $Q_0 = 1\,\text{GeV}$, $C_i = C_A = 3$, $n_b = 0$, $R = 0.6$, $A_{\text{tot}} = 4\pi$ and $A_{\text{Trans}} = 2\pi/3$ (the area of a single transverse region), corresponding to $L \simeq 1$ and $\mathcal{P} \simeq 0.35$. The result marked with a * is specific to the form of the soft toy model discussed in section 3.1.

one knows that shift as a function of rapidity, and hence one should determine $\langle\rho(y)\rangle$. The UE also affects jet energy *resolution*, through a term of the form $\sigma\sqrt{A_{\mathrm{jet}}}$. Thus we will also want to look at $\langle\sigma\rangle$ as a function of rapidity. A second way in which the UE affects jet energy resolution is that ρ itself is different event by event, so one might therefore examine its event-by-event distribution and its standard deviation S_d. The different ways in which UE affects jets was summarised in [12] with the following equation for the variance of the change in jet p_t due to the underlying event (neglecting back-reaction):

$$\langle\Delta p_{t,j}^2\rangle - \langle\Delta p_{t,j}\rangle^2 \simeq \langle\Sigma_{\mathrm{JA},R}^2\rangle\langle\rho\rangle^2 + \langle A_{\mathrm{JA},R}\rangle\langle\sigma^2\rangle + \langle A_{\mathrm{JA},R}\rangle^2 S_d^2 \,, \tag{47}$$

where $\langle A_{\mathrm{JA},R}\rangle$ is the average jet area and $\langle\Sigma_{\mathrm{JA},R}^2\rangle \sim R^4$ is the variance of the jet area. One sees that each of the terms involves a different characteristic of the UE: $\langle\rho\rangle$, $\langle\sigma\rangle$ and S_d^2.

Measurements of UE characteristics, as well as being of direct relevance to jet measurements, also have the power to constrain UE models. This has been the motivation for most UE studies to date and we believe that the range of UE characteristics discussed above would complement existing types of measurement and so provide an additional powerful set of constraints. Furthermore, as well as examining "local" quantities, such as $\langle\rho(y)\rangle$, one can also, for example, ask the question of whether there are long-range correlations between the magnitude of the UE in different parts of an event. In (prevalent) UE models that involve multiple independent $2 \to 2$ scatterings, one might expect these correlations to be modest, whereas in a BFKL-inspired model (as might derive from work like Ref. [16]), where one or more gluon ladders stretch across a whole event, one might expect them to be larger.

So far in this article we have treated the traditional and area/median based UE measurement approaches on a similar footing. The results of section 3 suggest that both provide reasonably adequate information about average characteristics of the underlying event: both methods introduced biases and different sources of biases cancelled partially, limiting their overall impact (though in the area/median approach the biases had much weaker dependence on the hard scale of the event). In contrast, as we saw in section 3.4 and its fig. 7, in the traditional approach the fluctuations due to hard perturbative emission were likely to dominate over fluctuations in the soft component of the UE, therefore for observables that are sensitive to fluctuations, we believe that the area/median approach is more robust. For this reason, in what follows, we shall concentrate on this second approach.

This section will be structured as follows: after outlining the set of Monte Carlo simulations that we use (section 4.1) and the event selection cuts (section 4.1), we examine to what extent the toy model is qualitatively similar to the realistic simulations (section 4.3). The motivation is that the toy model guides our intuition about the measurement procedure, and it is important to establish that this intuition is well founded. Having done so we then study (section 4.4) a selection of observables that are relevant for eq. (47), including their rapidity dependence and also their degree of correlation across different parts of a same event. This set of observables (together with some of those in section 4.3) would, we believe, be interesting to examine experimentally, both in terms of the direct information it provides for understanding the impact of UE on jets (and isolation, etc.) and in terms of its ability to constrain models.

4.1 Monte Carlo models used

We shall examine a series of Monte Carlo UE models: the UE that comes with the "old" virtuality-ordered shower in Pythia 6.4 [21], in the DW and DWT tunes by R. Field [18], identical at Tevatron energies, but with different energy-dependences; the UE that comes with

24

the "new" transverse-momentum-ordered shower in Pythia 6.4, in the S0A tune [5, 34, 35] by P. Skands; Herwig 6.5's [36, 37] default "soft" UE, which fails to reproduce various aspects of Tevatron data, but instructive for comparisons between different types of models; Herwig 6.5 with Jimmy 4.3 [4] in an ATLAS tune[13] by Moraes [18]. All models are based on multiple interactions except for Herwig's soft UE.

By default we use a pp centre of mass collision energy of $\sqrt{s} = 10\,\text{TeV}$, though we will also consider the energy dependence of some of the observables we study.

A comment is due on the fact that we will carry out our investigations at particle (hadron) level. Experiments may carry out measurements on tracks only (well measured, though with some low p_t threshold), on calorimeter information (subject to noise and noise-suppression thresholds) or on some combination of the the two. It is beyond the scope of this article to estimate the potentially substantial impact of detector effects on the results presented here. Nevertheless, we believe that the *differences* that we will see between various event-generator tunes should persist even after detector effects.

4.2 Event selection

We consider dijet events, where the leading jet, reconstructed with the anti-k_t algorithm [31] with $R = 0.6$ has $p_t > 100\,\text{GeV}$, the next hardest jet has $p_t > 80\,\text{GeV}$ and both jets are required to lie in the rapidity window $|y| < 4$. Note that since the cross section for jets falls steeply, a cut on jet p_t introduces a "trigger-bias", i.e. favours events where the UE is slightly larger than average.[14] The above cut is a default used for the study presented in this section. In some places we employ tighter cuts on the rapidities of the two hardest jets in order to study the impact of these jets on UE determination. Whenever this is the case, it will be indicated explicitly. The choice of the anti-k_t algorithm for event-selection here is motivated by the fact that it will be the first to be supported by ATLAS and CMS [39] and therefore will have been used by the experiments for their initial event selection.

For the determination of the properties of the UE within the jet-area/median approach, the anti-k_t algorithm is not suitable, as discussed in section 2.2. We will therefore use jets from the C/A algorithm with $R = 0.6$ for this task. All jet finding is performed with `FastJet 2.4.1` [40, 41], and we use the `ActiveAreaExplicitGhosts` option to calculate areas, because this ensures the safest treatment of pure ghost jets. The ghost area that is used is 0.01 and ghosts are placed up to $|y| < 9$ (we could, however, have used a smaller upper limit for most of the plots). The other parameters are left at their default values.

4.3 Comparisons of characteristics of MC and toy model

Given the importance of the toy model in guiding our understanding of the measurement of the underlying event, let us start by examining whether realistic Monte Carlo events bear any similarity to toy-model events. One way of doing this is to examine the R-dependence

---

[13]The non-default parameter setting are: PRSOF=0, JMRAD(73)=1.8, PTJIM= $2.8(\sqrt{s}/1.8)^{0.274}$, with CTEQ6L1 [38] PDFs (i.e. just the parameters in [18]).

[14] To investigate the impact of the trigger bias, we used the following procedure: we first measured the underlying event density ρ in a rapidity window $|\delta y| < 1$ around each hard jet, and then placed our hard jet cuts on $p_{t,\text{jet}} - \rho A_{\text{jet}}$ rather than on $p_{t,\text{jet}}$. The result for $\langle \rho \rangle$ in events that passed these cuts on "subtracted" jets came out about 10% lower than for events where we cut on unsubtracted jets. Since this is not too large an effect, and for reasons of simplicity, we have however chosen not to apply this procedure to our analysis as a whole.
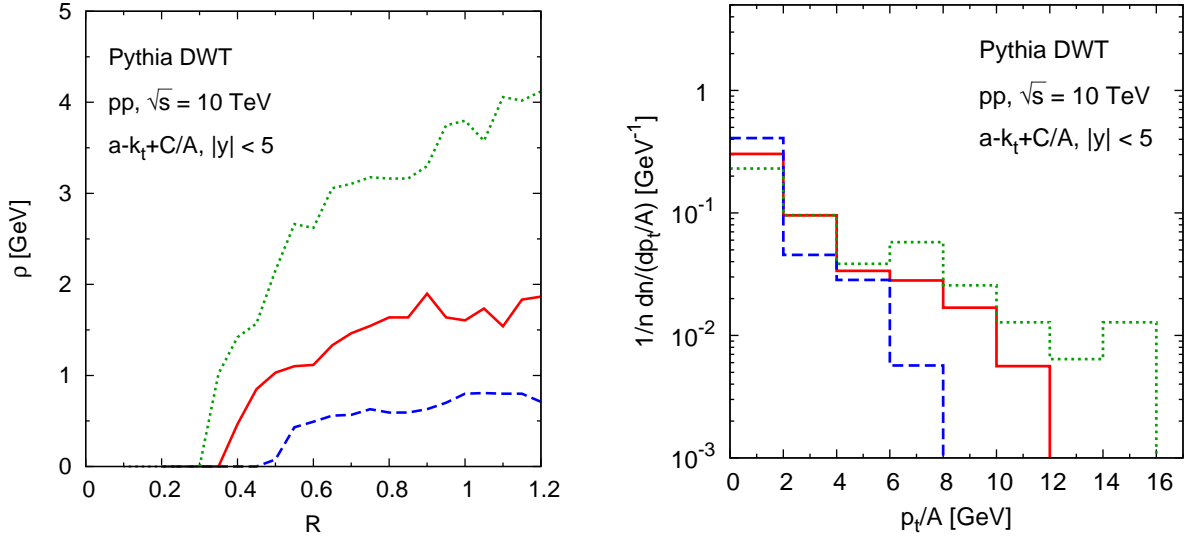
Figure 9: The left-hand plot shows $\rho(R)$ extracted with C/A jets in the area/median approach for 3 representative Pythia (DWT) events, which passed the selection cuts of section 4.2. The right-hand plot shows the corresponding histograms of p_t/A for the same 3 events.

of the extracted ρ, which, in the toy model, had a characteristic shape, fig. 5. Fig. 9 (left) shows $\rho(R)$ for three representative Pythia events (DWT tune). In each case one observes the turn-on at some threshold R value, followed by a roughly linear slope at larger R, precisely as expected. There is substantial variation in the curves from one event to the next, and one can trace this back to their distributions of p_t/A shown in fig. 9 (right): the blue (dashed) line which has small ρ and little R dependence corresponds to an event in which there is no activity at high p_t/A values. The green (dotted) curve, which has large ρ and strong R-dependence, has a correspondingly broad distribution of p_t/A values, with much activity at intermediate p_t values. In this event, the toy-model picture of a clean separation between soft and hard physics is somewhat challenged, though the general pattern of $\rho(R)$ having a turn-on followed by linear R-dependence still holds.

If we average $\rho(R)$ over many events we get fig. 10 (left), which shows results for several generators/tunes. One observation is that the Pythia tunes have a larger slope: based on the toy-model calculations, eq. (34), this can be interpreted as meaning that they have a larger value for the quantity σn_h, where n_h is the number of hard jets. The Herwig default curve, with no MPI, has the smallest slope (and ρ value). The average over many events for $\sigma(R)$ (normalised to $\langle\rho(R)\rangle$) is shown in fig. 10 (right). An interesting feature is the linear rise for $R = 0.2 - 0.4$ (and up to $R = 0.6$ for the Pythia tunes). In this region, one saturates the bound eq. (45), which implies that R is too small for a proper measurement of σ. For $R \gtrsim 0.6$ there is a shallower rise, which we interpret as being due to the presence of hard jets, as is the case in fig. 8 (though there is less curvature in the MC events than in the toy model).

To get an idea of the event-to-event variations of $\rho(R)$ we use the following procedure: given N events, for a given R, we sort the events according to $\rho(R)$. We then define the 10th percentile result for $\rho(R)$ to be the value of $\rho(R)$ for event $N/10$, and similarly for other percentiles. Fig. 11 shows the 10th, 20th, etc. percentile results for $\rho(R)$, as a function of R
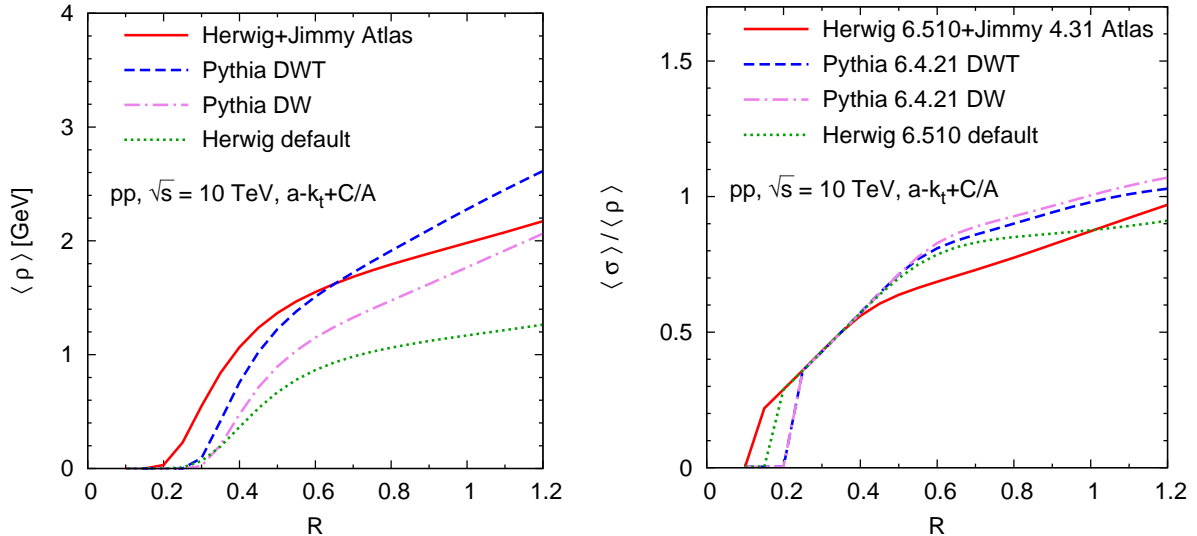
26

Figure 10: Average over many events of the extracted ρ (left) and σ (right), shown as function of R for four generators/tunes.

for our 4 generator/tune combinations (together with the average, for comparison).[15] One observes the sharp turn-on as a function of R (washed out in the $\langle\rho(R)\rangle$). The smaller the turn-on point, R_{crit}, the larger than value of ρ (and the larger the slope). The spread of events is noticeably large, both in the values of ρ and for R_{crit}, especially, for the latter, in the context of Herwig+Jimmy.

Though it is not our intention to highlight figures 10 and 11 as main results of this paper, we do note that it would be possible to make corresponding experimental measurements, and use them as input to MC tunes.

In the remaining parts of this section we shall concentrate on results extracted with $R = 0.6$, which, as anticipated in the toy-model section, seems to offer a reasonable compromise between being sufficiently large as to be well beyond the turn-on in most events, while not being too severely affected by the rise at large R that is induced by semi-hard radiation in the event.

4.4 Study of selected observables

4.4.1 Mean energy flow

Let us start by examining ρ and its rapidity dependence, fig. 12, for Tevatron and LHC energies.[16] The results for $\langle\rho\rangle$ in the central rapidity bin for Tevatron (left) should be strongly constrained by the standard UE measurements at Tevatron, and this is reflected in the small difference between S0A and DW, though there is a somewhat larger difference with the Herwig+Jimmy tune. The rapidity dependence is quite strong, with stronger suppression at

---

[15] A subtle point in the production of fig. 11 is that the event that provides the 10th percentile for (say) $R = 0.6$ is usually not the same event that provides it for $R = 0.8$. Thus the curves in fig. 11 are not the curves that would be obtained for individual events (these are far less smooth, cf. fig. 9), but can be thought of as some idealisation of these curves in a world free of fluctuations.

[16] To simplify the comparisons, we use the cuts of section 4.2 in both cases, though they involve a rapidity range that extends beyond the Tevatron's coverage.
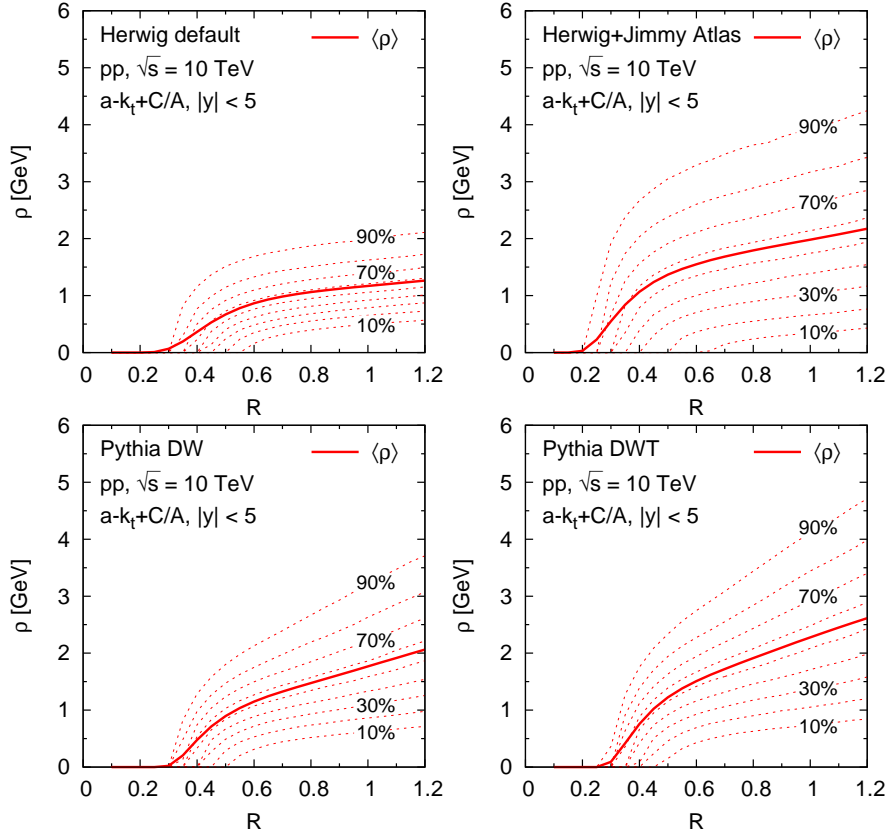
Figure 11: 10th, 20th, etc. percentiles for $\rho(R)$, as a function of R for four generator/tunes. Also shown is $\langle\rho(R)\rangle$ in each case.

forward rapidities for S0A and Herwig+Jimmy than for the DW tune. One should remember in examining the rapidity dependence that there are essentially no experimental constraints on the level of the underlying event at forward rapidities — it is therefore a model-dependent extrapolation. At LHC energies we see, fig. 12 (right), that differences appear between models also at central rapidities, reflecting an uncertainty in the extrapolation in energy. The DWT tune's energy-dependence is disfavoured based on RHIC [42] and lower-energy Tevatron data [25], but we include it to give an idea of the magnitude of possible differences.

One question that arises in the measurement of ρ is the possible bias from hard jets and the relation of this bias with the rapidity bin size or equivalently the area A_{tot} in which one measures ρ. In this context, recall that eq. (37) implies a bias from the Born jets that scales as n_b/A_{tot}. In fig. 12 the hard jets were free to lie anywhere within $|y| < 4$ and were included in the sample of jets used to obtain the median. In fig. 13 (left) the red (dark) dashed curve is the same as the DW result in fig. 12 (right). The red (dark) solid curve shows the impact of removing the 2 hardest jets from the median sample. One sees that this has a rather small effect. Next we examine an event sample in which the 2 hardest jets are in the central rapidity bin, so that that bin receives all the bias from the Born jets. If those jets are included in the set used to calculate the median (dotted green histogram), then the impact on $\langle\rho(y)\rangle$ in that bin becomes noticeable, $\mathcal{O}(30\%)$. Removing them from the set for the median (solid green curve) brings us almost back to the basic "all-2" result for the full dijet sample. That the
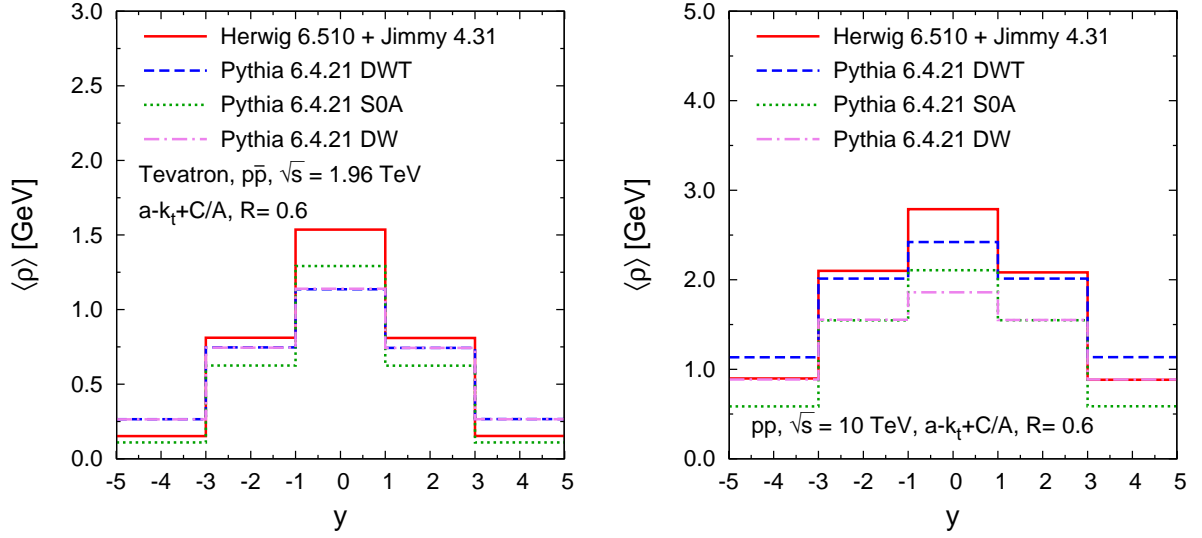
Figure 12:   The rapidity dependence of the average underlying event activity, $\langle\rho(y)\rangle$, shown for $\sqrt{s} = 1.96\,\text{TeV}$ Tevatron ($p\bar{p}$) simulations (left) and $\sqrt{s} = 10\,\text{TeV}$ LHC (pp) simulations (right).
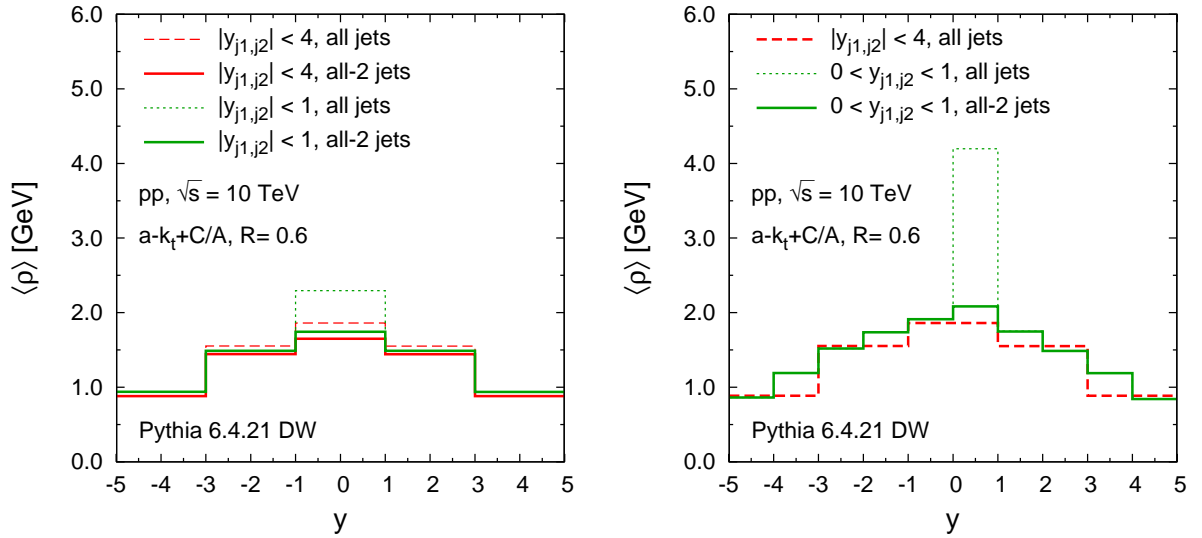

Figure 13:   Dependence of the $\langle\rho(y)\rangle$ results on the rapidities of the hard jets and on the choice of whether to exclude the two hardest jets from the median procedure. The left-hand plot shows this for our standard rapidity-bin size, $\delta y = 2$, corresponding to $A_{\text{tot}} = 4\pi$, while the right-hand plot shows what happens if we choose smaller bins, $\delta y = 1$, i.e.$= A_{\text{tot}} = 2\pi$. The results correspond to the Pythia DW tune.
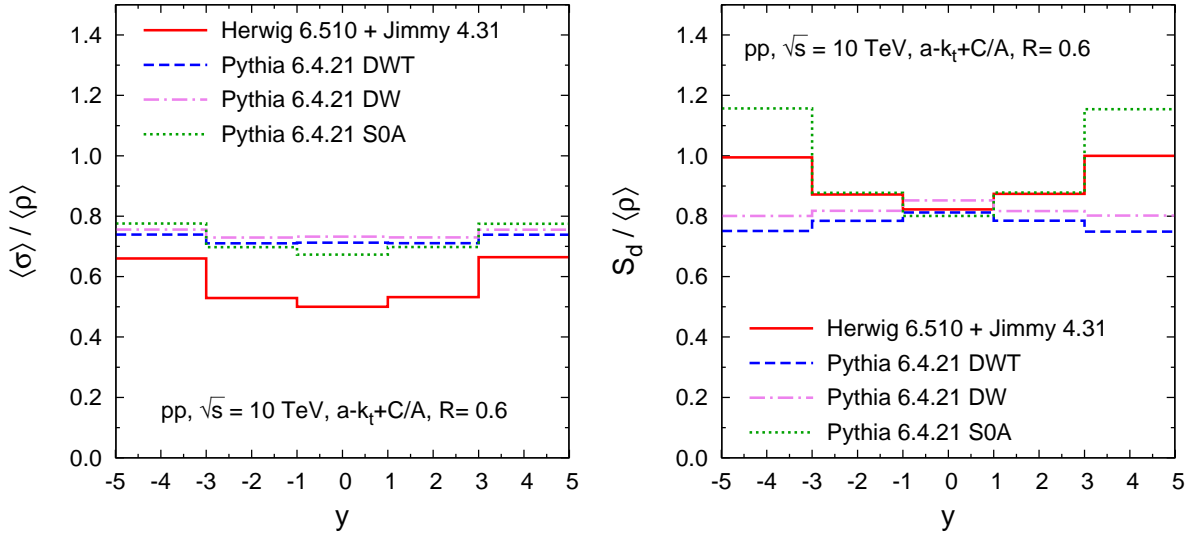
Figure 14: Left: $\langle\sigma\rangle/\langle\rho\rangle$ as a function of rapidity. Right: $S_d/\langle\rho\rangle$ as a function of rapidity.

result is not fully identical is a consequence of the fact that when the two hard jets are central there is an increased probability that the 3rd hardest jet will also be central, thus biasing very slightly the central-y bin.

The right-hand plot of fig. 13 shows what happens if we reduce the rapidity bin size, causing ρ to be measured in regions of area $A_{tot} = 2\pi$ rather than $A_{tot} = 4\pi$. Since the impact of the Born particles is inversely proportional to A_{tot}, requiring the Born particles to be in the central bin has a noticeably larger effect for the smaller rapidity bin size.[17] Discarding the two hard jets brings us back to a result that is roughly in accord with that for the larger bin size.

The conclusion from fig. 13 is that if one's event selection does not constrain the hard jets to be in the same bin as that used for measuring ρ and if the bin area is sufficiently large, $A_{tot} \gtrsim 12$, then biases from the hard jets are quite small. In what follows, we will normally use $A_{tot} = 4\pi$ and leave in the hard jets, in order to keep the analysis as simple as possible.

4.4.2 Fluctuations

In fig. 14 we examine fluctuations, now only for LHC. The left-hand plot shows the size of intra-event fluctuations, through the ratio of $\langle\sigma(y)\rangle$ to $\langle\rho(y)\rangle$, while the right hand plot shows inter-event fluctuations, through $S_d(y)/\langle\rho(y)\rangle$. Fluctuations have not been as directly tuned as energy flow. Despite this the intra-event fluctuations are very similar across all the Pythia tunes and almost independent of rapidity when normalised to $\langle\rho(y)\rangle$. Herwig+Jimmy's intra-event fluctuations are somewhat smaller, but do have rapidity dependence.

Concerning inter-event fluctuations, fig. 14 (right), the two virtuality ordered (DW/DWT) Pythia models are again flat, whereas the p_t-ordered shower has increasing fluctuations at forward rapidities. Herwig is intermediate between the two sets of Pythia results.

One observation is that whereas DW/DWT have almost identical intra and inter-event fluctuations, Herwig's intra-event fluctuations are nearly 40% smaller than the inter-event

---

[17]The effect is definitely larger than would be expected based on eq. (34), perhaps a reflection of the non-Gaussianity of the distribution of p_{tj}/A_j for the events under consideration.
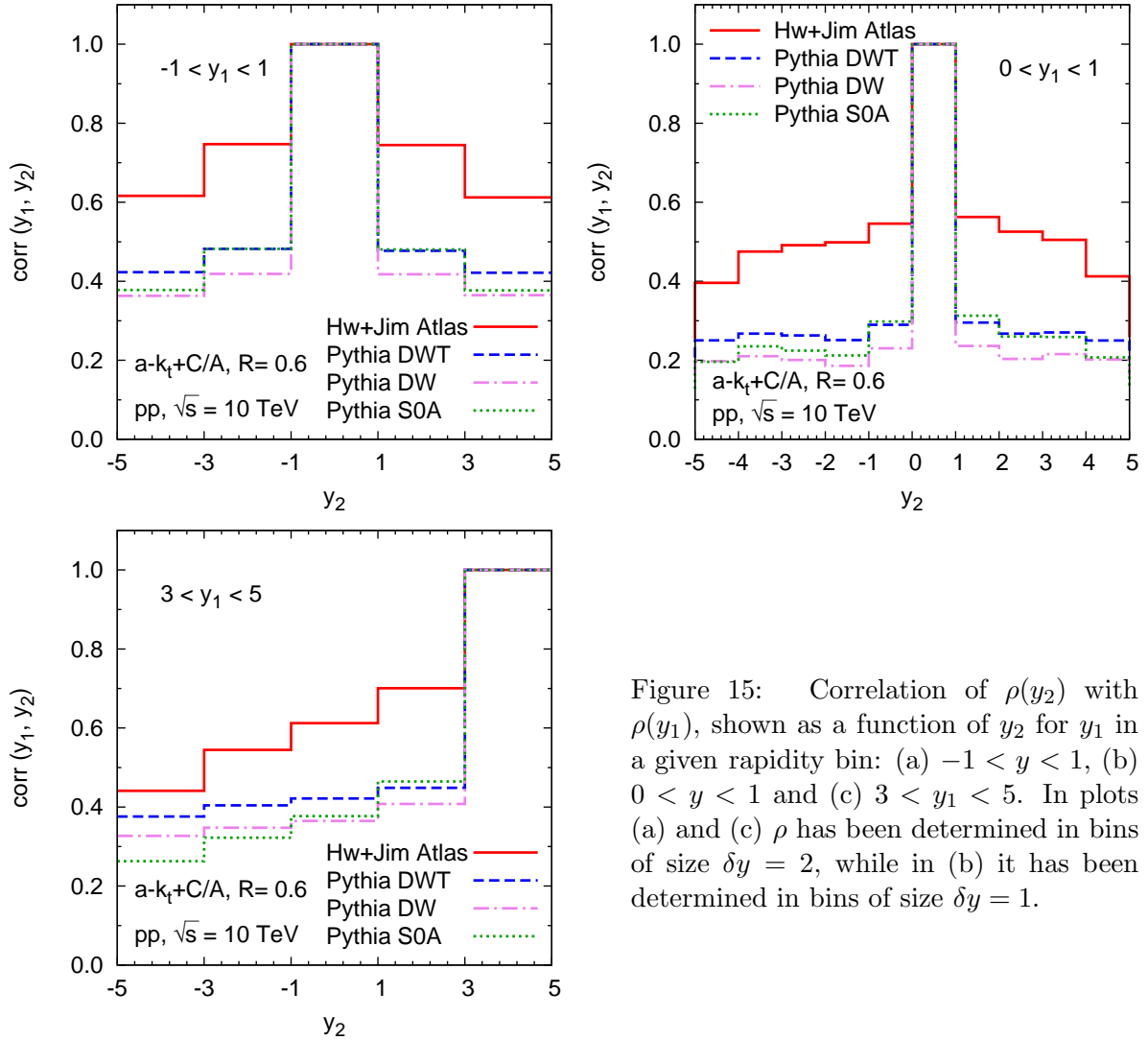
Figure 15: Correlation of $\rho(y_2)$ with $\rho(y_1)$, shown as a function of y_2 for y_1 in a given rapidity bin: (a) $-1 < y < 1$, (b) $0 < y < 1$ and (c) $3 < y_1 < 5$. In plots (a) and (c) ρ has been determined in bins of size $\delta y = 2$, while in (b) it has been determined in bins of size $\delta y = 1$.

fluctuations. This is reflected also when we examine correlations between ρ in different parts of the event,

$$\mathrm{corr}(y_1, y_2) = \frac{\langle \rho(y_1)\rho(y_2) \rangle - \langle \rho(y_1) \rangle \langle \rho(y_2) \rangle}{S_d(y_1)S_d(y_2)} \tag{48}$$

as shown in fig. 15. The correlations are noticeably larger for Herwig+Jimmy than they are for all the Pythia tunes. In determining the correlations it was important that we used sufficiently large rapidity bins. Comparing the upper-left plot from fig. 15 ($\delta y = 2$) and the upper-right plot ($\delta y = 1$) one sees that the smaller rapidity bins lead to noticeably smaller measured correlations. We interpret this as follows: in small rapidity bins, the "statistics" of jets for measuring the ρ value are more limited. This increases the error on the determination of ρ, thus reducing the maximum amount of correlation that can be observed between different bins.

The final quantity that we examine is the event-by-event distribution of ρ, fig. 16, for a central rapidity bin (left) and a forward rapidity bin (right). Perhaps the most striking characteristic of these plots is the very broad nature of these distributions, which are far
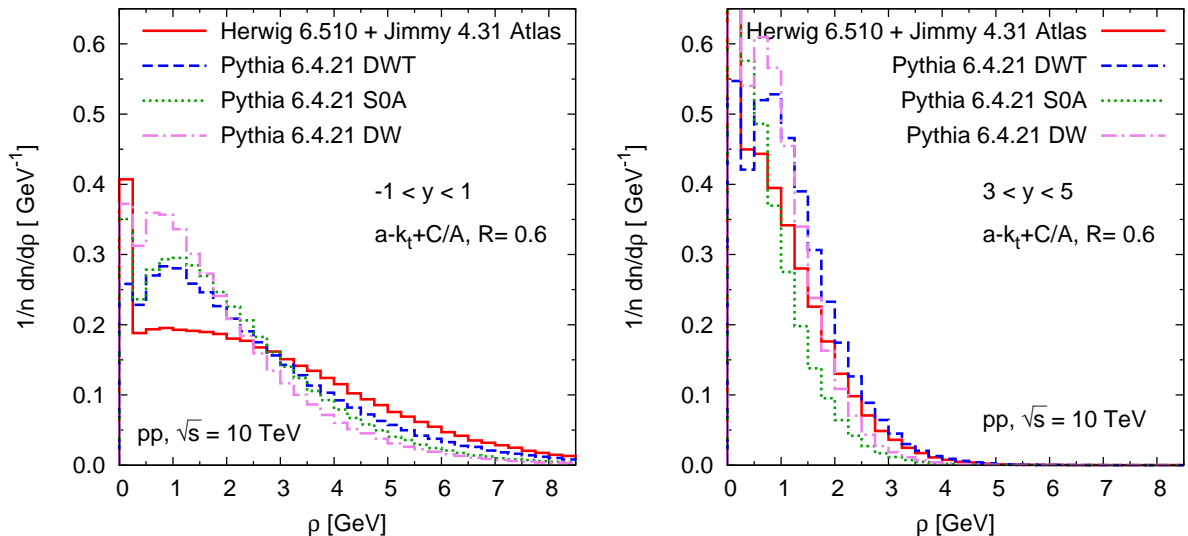
Figure 16: The event-by-event distribution of ρ for a selection of generators, in a central rapidity bin, $|y| < 1$ (left-hand plot) and a forward rapidity bin, $3 < y < 5$ (right).

from being Gaussian distributions of width S_d centred on $\langle\rho\rangle$. The right-hand plot also has a significant bin at $\rho = 0$ — i.e. there is a substantial number of events for which at least half of the jets at forward rapidities are pure ghost jets.

4.4.3 Energy dependence of results

Fig. 17 summarises the energy-dependence of the average energy flow $\langle\rho\rangle$ (left) and of the intra-event fluctuations $\langle\sigma\rangle$ normalised to $\langle\rho\rangle$ (right) in the central rapidity bin $|y| < 1$. The features of note are that for $\langle\rho\rangle$, the Herwig+Jimmy Atlas tune has a significantly steeper energy dependence than the DW and S0A Pythia tunes, somewhat more like DWT. For $\langle\sigma\rangle/\langle\rho\rangle$, the Pythia tunes have almost no energy-dependence, whereas Herwig has substantial energy-dependence. We have not shown the energy dependence of $S_d/\langle\rho\rangle$ because it is essentially independent of $\sqrt{s}$ for all generators.

5 Conclusions

The distinction between underlying event and the "hard" part of hadron-collider events is ill-defined in QCD. Nevertheless, physically, one may picture the UE as being the component of a hadron-hadron scattering that fills the event fairly uniformly with low-p_t radiation. A goal of this article was to investigate how different UE-measurement strategies fare in separating such a low-p_t component from the hard part of the event.

To do so, we developed a simple toy-model for events, with two distinct components, one soft, corresponding to the UE, the other perturbative and hard. Within this model it is quite straightforward to establish to what extent a given UE measurement strategy correctly extracts just the soft component.

The two UE measurement strategies that we investigated are the "traditional" approach, measuring radiation in regions transverse to a leading jet, and the "jet-area/median" approach. Both involve strategies to help separate out the soft and hard components: the use of
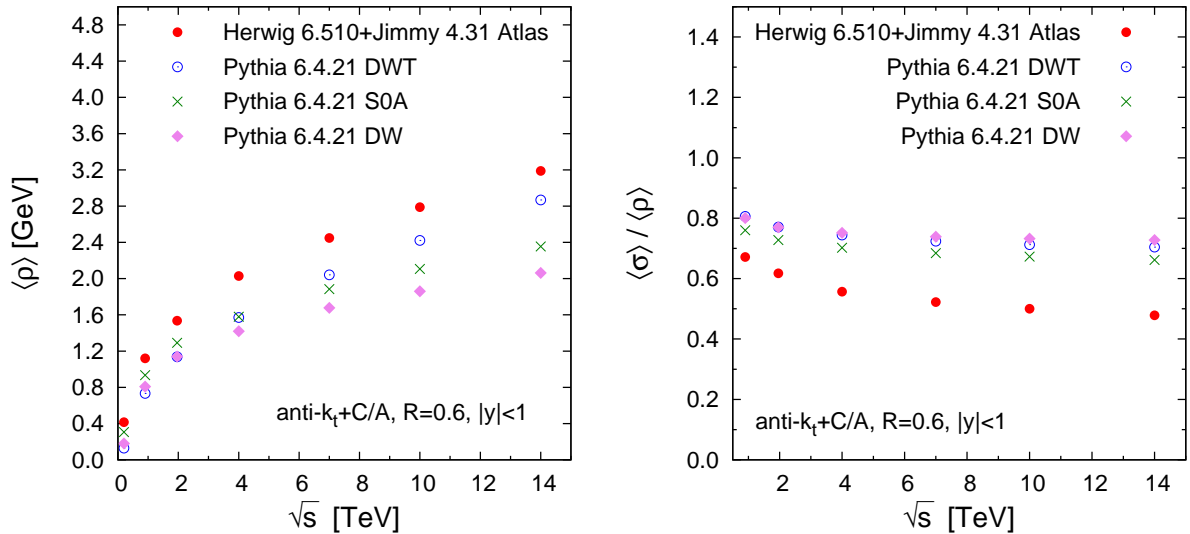
Figure 17: Level of the underlying event $\langle\rho\rangle$ (left) and its intra-event fluctuations $\langle\sigma\rangle/\langle\rho\rangle$ (right) in central rapidity rapidity bin $|y| < 1$ as a function of energy of pp collision. The points correspond to the energies of RHIC, Tevatron and a set of possible LHC $\sqrt{s}$ values.

the Av/Min/Max transverse regions in the traditional approach, the use of median activity rather than average in the jet-area/median approach. One result from the toy model is a quantification of how those strategies fare in the extraction of its soft component. A second result is a determination of the nature of the residual effects due to perturbative (hard) radiation. These two results could be expressed analytically in terms of the parameters of the measurement procedures (transverse-region area, jet radius, providing useful guidance in choosing them) and of characteristics of the hard scattering (notably the value of the hard scale and the properties of the soft component). We also examined the question of event-to-event fluctuations in the extraction of the characteristics of the soft component. Our toy-model results are summarised in section 3.6.

Practically, one conclusion from this work is that for determinations of averaged quantities, for example the mean transverse-momentum density per unit area, $\langle\rho\rangle$, both the TransMin and the area/median measurement methods give a fair determination of the soft component, as long as the momentum transfer of the hard scattering is not too large ($\lesssim 100\,\text{GeV}$ for the LHC; for higher momentum transfers, the TransMin method is affected by a rapidly increasing hard contribution). In particular, for the parameter choices used or advocated in the literature, the two kinds of bias seen in the toy model, mismeasurement of the soft component and contamination from the hard component, tend to partially cancel each other, giving a limited overall bias, of order 20%. In contrast, for event-by-event measurements and determinations of fluctuations of the soft component, the traditional approach is significantly affected by rare "outliers", cf. table 1.

For this reason, in the full Monte Carlo studies of section 4 we concentrated on the area/median approach. The results included a validation of the main qualitative prediction from the toy model, namely the structure of the R-dependence of the extracted ρ, section 4.3. Section 4.4 showed a range of possible observables whose measure we advocate at LHC: rapidity dependence of the UE, nature of the event-to-event fluctuations, and intra-event fluctuations and correlations. Though existing measurements may indirectly constrain some

33

of these features of the UE, we believe that they are of sufficient practical importance that they deserve dedicated measurements, especially as they differ noticeably between various Monte Carlo models.

Acknowledgements

A Toy model UE calculations

A.1 Threshold and asymptotic regions

With the exponential model for the single-particle p_t distribution, eq. (4), there are two asymptotic limits of interest: the threshold for $\nu A \to \ln 2$ and the asymptotic $\nu A \to \infty$ region (for brevity here we write $A \equiv A_{\text{tile}}$). Writing eq. (10) as an expansion around $\nu A = \ln 2$, one obtains

$$\frac{1}{2}\left(1 + \ln 2 - \nu A\right) + \frac{\ln 2}{2}\int_0^{A\langle\rho_{\text{ext}}\rangle}\frac{dp_t}{\mu}e^{-p_t/\mu} + \mathcal{O}\left((\nu A - \ln 2)^2\right) = \frac{1}{2}, \tag{49}$$

where we have used eq. (6a) for dP/dp_t and kept only the δ-function term and the first term of the sum over n. To first order in $\nu A - \ln 2$, this gives us the behaviour in the turn-on region,

$$A\langle\rho_{\text{ext}}\rangle = \frac{\nu A - \ln 2}{\ln 2}\mu + \mathcal{O}\left((\nu A - \ln 2)^2\mu\right), \tag{50}$$

or, equivalently, using $\mu = \rho/\nu \simeq \rho A/\ln 2$ for $\nu A \simeq \ln 2$,

$$\frac{\langle\rho_{\text{ext}}\rangle}{\rho} = \frac{\nu A - \ln 2}{\ln^2 2} + \mathcal{O}\left((\nu A - \ln 2)^2\right). \tag{51}$$

The approximation eq. (11) reproduces the linear dependence on $\nu A - \ln 2$ at threshold, though its slope there, $d\langle\rho_{\text{ext}}\rangle/d(\nu A) = 2$, differs slightly from the exact slope of $1/\ln^2 2$.

At large νA, by examining the numerical solutions to eq. (10), we have determined the following relation for the asymptotic behaviour of ρ_{ext},

$$\langle\rho_{\text{ext}}\rangle = \nu\mu - \frac{\mu}{2A} + \mathcal{O}\left(\frac{\mu}{\nu A^2}\right) = \rho - \frac{\rho}{2\nu A} + \mathcal{O}\left(\frac{\rho}{\nu^2 A^2}\right). \tag{52}$$

This is reproduced by the approximation of eq. (11).

An approximation that is closer still to the full result, with the correct coefficients in both limits, is

$$\rho_{\text{ext}} = \rho\frac{X + X^2}{(\ln 2)^2 + X^2 + \frac{3}{2}X}, \qquad X = \nu A - \ln 2, \tag{53}$$

however the difference between this and eq. (11) is irrelevant for all practical purposes.

A.2 Variant of toy model

We can also consider a model with

$$\frac{1}{P_1}\frac{dP_1}{dp_t} = \frac{4p_t}{\mu^2}e^{-2p_t/\mu}\,. \tag{54}$$

and correspondingly

$$\frac{1}{P_n}\frac{dP_n}{dp_t} = \frac{1}{\mu}\frac{2^{2n}}{(2n-1)!}\left(\frac{p_t}{\mu}\right)^{2n-1}e^{-2p_t/\mu}\,. \tag{55}$$

This model has the property that $\sigma = \sqrt{3\nu/2}\mu$. For ν near $\ln 2$ this leads to

$$\frac{1}{2}(1-X) + \frac{\ln 2}{2}\int_0^{A\langle\rho_{\mathrm{ext}}\rangle}dp_t\,\frac{4p_t}{\mu^2}e^{-2p_t} + \mathcal{O}\left(X^2\right) = \frac{1}{2}\,, \tag{56}$$

with X defined as in eq. (53), resulting in

$$A\langle\rho_{\mathrm{ext}}\rangle = \mu\sqrt{\frac{\nu A - \ln 2}{2\ln 2}} + \mathcal{O}\left((\nu A - \ln 2)^{3/2}\mu\right)\,, \tag{57}$$

or, equivalently,

$$\langle\rho_{\mathrm{ext}}\rangle = \frac{\rho}{\ln 2}\sqrt{\frac{\nu A - \ln 2}{2\ln 2}} + \mathcal{O}\left((\nu A - \ln 2)^{3/2}\rho\right)\,. \tag{58}$$

At large ν one finds, again numerically,

$$\langle\rho_{\mathrm{ext}}\rangle = \nu\mu - \frac{\mu}{3A} + \mathcal{O}\left(\frac{\mu}{\nu A^2}\right) = \rho - \frac{\rho}{3\nu A} + \mathcal{O}\left(\frac{\rho}{\nu^2 A^2}\right)\,. \tag{59}$$

The two above equations both imply that $\langle\rho_{\mathrm{ext}}\rangle$ turns on and approaches its asymptotic value somewhat faster than in the model with $\frac{dP_1}{dp_t} \propto e^{-p_t/\mu}/\mu$. A reasonable analytic approximation for $\langle\rho_{\mathrm{ext}}\rangle$ over the whole domain is

$$\langle\rho_{\mathrm{ext}}\rangle \simeq \rho\sqrt{\frac{\nu A - \ln 2}{\nu A - \ln 2 + \frac{2}{3}}}\,\Theta(\nu A - \ln 2)\,, \tag{60}$$

which has the correct large-ν behaviour, and very nearly the correct coefficient for the $\sqrt{\nu A - \ln 2}$ turn-on.

B Fluctuations in area/median extraction of ρ

B.1 Pure soft case

To determine the event-to-event fluctuations S_d in the area/median extraction of ρ when the intrinsic event-to-event fluctuations are zero, it is convenient to work in the limit $\nu A_{\mathrm{jet}} \ll 1$ such that the probability distribution of ρ_{jet} is close to a Gaussian,[18]

$$\frac{dP_{\mathrm{jet}}}{d\rho}(\delta\rho_{\mathrm{jet}}) \equiv \frac{1}{N}\frac{dN}{d\rho_{\mathrm{jet}}} = \frac{1}{\sigma}\sqrt{\frac{\langle A_{\mathrm{jet}}\rangle}{2\pi}}\exp\left(-\frac{\langle A_{\mathrm{jet}}\rangle}{2\sigma^2}\delta\rho_{\mathrm{jet}}^2\right)\,, \qquad \delta\rho_{\mathrm{jet}} \equiv \rho_{\mathrm{jet}} - \rho\,. \tag{61}$$

---

[18]We will ignore the issue of fluctuations in A_{jet} itself, i.e. it is really the tiled case that we consider here rather than the full jet case. Nevertheless the good agreement that we find with numerical studies in section 3.4 suggests that it is not illegitimate to ignore jet area fluctuations for our purposes here.

The corresponding cumulative probability distribution for the jets is given by

$$P_{\text{jet}}(\delta\rho) = \int_{-\infty}^{\delta\rho+\rho} d\rho' \frac{1}{N}\frac{dN}{d\rho} = \frac{1}{2}\left(1 + \text{Erf}\left(\frac{\delta\rho}{\sigma}\sqrt{\frac{\langle A_{\text{jet}}\rangle}{2}}\right)\right). \tag{62}$$

If the number N of jets is odd, $N = 2m + 1$, then the probability distribution of the median $\delta\rho$ is obtained from the product of the probability of having one jet with $\rho_{\text{jet}} = \rho - \delta\rho$, m jets with ρ smaller than this, and m jets with ρ larger than this:

$$\frac{dP_{\text{med}}}{d\rho}(\delta\rho) = (2m+1)\frac{(2m)!}{(m!)^2}\frac{dP_{\text{jet}}}{d\rho}(\delta\rho)\,[P_{\text{jet}}(\delta\rho)]^m\,[1 - P_{\text{jet}}(\delta\rho)]^m. \tag{63}$$

Making use of the expansion of the error function

$$P_{\text{jet}}(\delta\rho) = \frac{1}{2} + \sqrt{\frac{\langle A_{\text{jet}}\rangle}{2\pi}}\frac{\delta\rho}{\sigma} + \mathcal{O}\left(\left(\frac{\delta\rho}{\sigma}\right)^3\right), \tag{64}$$

working in the large m and small $\delta\rho/\sigma$ limit, and making use also of Stirling's formula $m! \simeq \sqrt{2\pi m}(m/e)^m$, one can approximate eq. (63) as

$$\frac{dP_{\text{med}}}{d\rho}(\delta\rho) \simeq \frac{\sqrt{2m\langle A_{\text{jet}}\rangle}}{\pi\sigma}\left(1 - \frac{2\langle A_{\text{jet}}\rangle}{\pi}\frac{\delta\rho^2}{\sigma^2}\right)^m \simeq \frac{\sqrt{N\langle A_{\text{jet}}\rangle}}{\pi\sigma}\exp\left(-\frac{N\langle A_{\text{jet}}\rangle}{\pi}\frac{\delta\rho^2}{\sigma^2}\right), \tag{65}$$

where in the last step we have replaced $m \simeq N/2$. Using the relation $N\langle A_{\text{jet}}\rangle = A_{\text{tot}}$, we finally obtain the following result for the standard deviation of the extracted median ρ values,

$$S_{d,\text{med}}^{(\text{soft})} \simeq \sigma\sqrt{\frac{\pi}{2A_{\text{tot}}}}. \tag{66}$$

This is about 25% larger than the standard deviation that would be obtained for ρ extracted as an average of the ρ_{jet} values over the same total area. This moderate enhancement of fluctuations in the pure soft case is part of the price that one pays in exchange for the median's greater resilience to hard contamination.

B.2   Hard contamination

The result eq. (34) for the average discrepancy in ρ_{ext} due to hard contamination can be obtained in an alternative manner, which will be more useful for estimating fluctuations. In this approach we imagine some distribution of soft jets, and then add in the hard partons. Some of the hard partons will enter jets whose ρ_{jet} is already above the median value for ρ. These hard partons will not affected ρ_{ext}. The remaining hard partons (a number k) will enter jets that were below the median. These jets will acquire much larger transverse momenta, taking them well above the median. Thus it becomes necessary to recalculate the median, which will be shifted by k soft jets' worth. From eq. (61), and working in the large-N limit (throughout this section), this translates to an average shift in ρ_{ext} of

$$\langle\delta\rho\rangle = k\delta_1, \qquad \delta_1 \equiv \frac{\sigma}{N}\sqrt{\frac{2\pi}{\langle A_{\text{jet}}\rangle}}, \tag{67}$$

where δ_1 is the 1-jet shift. Substituting $\langle k \rangle = \langle n_h \rangle /2$ gives

$$\langle \delta \rho \rangle = \frac{\langle n_h \rangle}{2} \delta_1 = \sigma \sqrt{\frac{\pi}{2 \langle A_{\text{jet}} \rangle} \frac{\langle n_h \rangle}{N}} \, , \tag{68}$$

in accord with the first order term in $\langle n_h \rangle /N$ in eq. (33).

We will consider two main sources of fluctuations in this result. Let us first imagine that $k = 1$. The median will shift up by one jet, and the distribution of $\delta \rho$ will be simply be given by the distribution of the difference in ρ_{jet} between two neighbouring jets in the sorted sequence of jets (at position in the sequence that is near the median). That distribution is an exponential distribution with mean δ_1,

$$\frac{dP}{d\delta \rho}(k = 1) = \frac{1}{\delta_1} \exp\left[-\delta \rho / \delta_1\right] \, . \tag{69}$$

The distribution of the shift for k jets is

$$\frac{dP}{d\delta \rho}(k) = \frac{(\delta \rho)^{k-1}}{k! \, (\delta_1)^k} \exp\left[-\delta \rho / \delta_1\right] \, , \tag{70}$$

with the standard deviation $\sqrt{k}\delta_1$.

The second source of fluctuations comes from the fact that k is not constant but rather has a Poisson distribution with mean $\langle n_h \rangle /2$ and standard deviation $\sqrt{\langle n_h \rangle /2}$.[19] Combining this with the fluctuation on ρ for fixed k, leads to an overall standard deviation of

$$S_{d,\text{med}}^{(\text{hard})} = \sqrt{\langle n_h \rangle}\delta_1 = \sigma \sqrt{\frac{2\pi}{\langle A_{\text{jet}} \rangle}} \left(\frac{\sqrt{\langle n_h \rangle}}{N} + \mathcal{O}\left(\frac{\langle n_h \rangle^{3/2}}{N^2} \right) \right) \, , \tag{71}$$

In terms of R and A_{tot}, this becomes

$$S_{d,\text{med}}^{(\text{hard})} = \sigma R \sqrt{2\pi c_J} \left(\frac{\sqrt{\langle n_h \rangle}}{A_{\text{tot}}} + \mathcal{O}\left(\frac{\langle n_h \rangle^{3/2} c_J R^2}{A_{\text{tot}}^2} \right) \right) \tag{72a}$$

$$\simeq 1.79 \frac{\sigma R}{\sqrt{A_{\text{tot}}}} \left(\frac{C_i}{C_A} L + 4.0 \frac{n_b}{A_{\text{tot}}} \right)^{\frac{1}{2}} + \cdots \tag{72b}$$

where $L \simeq 1$ was defined in eq. (35).

References

[1] A. Abulencia *et al.* [CDF - Run II Collaboration], Phys. Rev. D **75** (2007) 092006 [Erratum-ibid. D **75** (2007) 119901] [arXiv:hep-ex/0701051].

[2] R. Harnik and T. Wizansky, arXiv:0810.3948 [hep-ph].

[3] T. Sjostrand and M. van Zijl, Phys. Rev. D **36** (1987) 2019.

---

[19]The statement about the Poisson distribution holds for the perturbatively radiated partons, but not, strictly speaking, for the Born partons. Nevertheless, for simplicity, we shall treat the Born and perturbatively radiated partons in common here.

[4] J. M. Butterworth, J. R. Forshaw and M. H. Seymour, Z. Phys. C **72** (1996) 637 [arXiv:hep-ph/9601371].

[5] T. Sjostrand and P. Z. Skands, Eur. Phys. J. C **39** (2005) 129 [arXiv:hep-ph/0408302].

[6] T. Sjostrand, S. Mrenna and P. Skands, Comput. Phys. Commun. **178** (2008) 852 [arXiv:0710.3820 [hep-ph]].

[7] M. Bahr *et al.*, arXiv:0812.0529 [hep-ph].

[8] M. Bahr, S. Gieseke and M. H. Seymour, JHEP **0807** (2008) 076 [arXiv:0803.3633 [hep-ph]].

[9] T. Gleisberg, S. Hoche, F. Krauss, M. Schonherr, S. Schumann, F. Siegert and J. Winter, JHEP **0902** (2009) 007 [arXiv:0811.4622 [hep-ph]].

[10] R. Corke and T. Sjostrand, arXiv:0911.1909 [hep-ph].

[11] M. Dasgupta, L. Magnea and G. P. Salam, JHEP **0802** (2008) 055 [arXiv:0712.3014 [hep-ph]].

[12] M. Cacciari, G. P. Salam and G. Soyez, JHEP **0804** (2008) 005 [arXiv:0802.1188 [hep-ph]].

[13] M. Cacciari and G. P. Salam, Phys. Lett. B **659** (2008) 119 [arXiv:0707.1378 [hep-ph]].

[14] J. R. Gaunt and W. J. Stirling, arXiv:0910.4347 [hep-ph].

[15] Z. Nagy and D. E. Soper, Phys. Rev. D **74** (2006) 093006 [arXiv:hep-ph/0610028].

[16] E. Avsar, G. Gustafson and L. Lonnblad, JHEP **0701** (2007) 012 [arXiv:hep-ph/0610157]; E. Avsar, G. Gustafson and L. Lonnblad, JHEP **0712** (2007) 012 [arXiv:0709.1368 [hep-ph]].

[17] T. Sjostrand and P. Z. Skands, JHEP **0403** (2004) 053 [arXiv:hep-ph/0402078].

[18] M. G. Albrow *et al.* [TeV4LHC QCD Working Group], arXiv:hep-ph/0610012.

[19] D. Kar and R. Field (on behalf of the CDF collaboration), CDF/PUB/CDF/PUBLIC/9531, July 2008.3

[20] S. Salur, arXiv:0910.2081 [hep-ex], and references therein.

[21] T. Sjostrand, S. Mrenna and P. Skands, JHEP **0605** (2006) 026 [arXiv:hep-ph/0603175].

[22] G. Marchesini, B. R. Webber, G. Abbiendi, I. G. Knowles, M. H. Seymour and L. Stanco, Comput. Phys. Commun. **67** (1992) 465; G. Corcella *et al.*, JHEP **0101** (2001) 010 [hep-ph/0011363].

[23] G. Marchesini and B. R. Webber, Phys. Rev. D **38** (1988) 3419.

[24] C. Albajar *et al.* [UA1 Collaboration], CERN-EP/88-29, Nucl. Phys. B **309** (1988) 405.

[25] D. E. Acosta *et al.* [CDF Collaboration], Phys. Rev. D **70** (2004) 072002 [arXiv:hep-ex/0404004].

[26] Y. L. Dokshitzer, G. D. Leder, S. Moretti and B. R. Webber, JHEP **9708**, 001 (1997) [hep-ph/9707323];

[27] M. Wobisch and T. Wengler, arXiv:hep-ph/9907280; M. Wobisch, positron proton collisions at $\sqrt{s} = 300$-GeV," DESY-THESIS-2000-049.

[28] S. Catani, Y. L. Dokshitzer, M. H. Seymour and B. R. Webber, collisions," Nucl. Phys. B **406**, 187 (1993).

[29] S. D. Ellis and D. E. Soper, Phys. Rev. D **48**, 3160 (1993) [hep-ph/9305266].

[30] S. Salur [STAR Collaboration], Eur. Phys. J. C **61** (2009) 761 [arXiv:0809.1609 [nucl-ex]].

[31] M. Cacciari, G. P. Salam and G. Soyez, JHEP **0804** (2008) 063 [arXiv:0802.1189 [hep-ph]].

[32] G. P. Salam and G. Soyez, JHEP **0705** (2007) 086 [arXiv:0704.0292 [hep-ph]].

[33] Z. Nagy, order," Phys. Rev. Lett. **88** (2002) 122003 [hep-ph/0110315]; hadron collision," Phys. Rev. D **68** (2003) 094002 [hep-ph/0307268].

[34] P. Skands and D. Wicke, Eur. Phys. J. C **52** (2007) 133 [arXiv:hep-ph/0703081].

[35] C. Buttar *et al.*, arXiv:hep-ph/0604120.

[36] G. Corcella *et al.*, JHEP **0101** (2001) 010 [arXiv:hep-ph/0011363].

[37] G. Corcella *et al.*, arXiv:hep-ph/0210213.

[38] J. Pumplin, D. R. Stump, J. Huston, H. L. Lai, P. M. Nadolsky and W. K. Tung, JHEP **0207** (2002) 012 [arXiv:hep-ph/0201195].

[39] S. Resconi (for the ATLAS collaboration), presentation at the Hadron Collider Physics Symposium, Evian, France, November 2009;
V. D. Elvira (for the CMS collaboration), presentation at the Hadron Collider Physics Symposium, Evian, France, November 2009.

[40] M. Cacciari and G. P. Salam, Phys. Lett. B **641** (2006) 57 [arXiv:hep-ph/0512210].

[41] M. Cacciari, G. P. Salam and G. Soyez, FastJet, http://fastjet.fr

[42] G. Webb for the STAR Collaboration, presented at the APS April meeting, Denver, Colorado, May 2009, `http://drupal.star.bnl.gov/STAR/files/APS_WEBB_UE_0.pdf`